

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M nucleic - nucleic search, using sw model

Run on: September 11, 2004, 10:05:53 ; Search time 14891.6 Seconds
(without alignment)
7040.653 Million cell updates/sec

Title: US-09-939-408a-12

Perfect score: 2419

Sequence: 1 ctccgcctcactcctaaccc.....gagttcaaaaaaaaaa 2419

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb ba.*
2: gb hcg.*
3: gb in.*
4: gb om.*
5: gb ov.*
6: gb pat.*
7: gb ph.*
8: gb pl.*
9: gb pr.*
10: gb ro.*
11: gb scs.*
12: gb sy.*
13: gb un.*
14: gb vi.*
15: em ba.*
16: em fun.*
17: em hum.*
18: em in.*
19: em ml.*
20: em om.*
21: em or.*
22: em ov.*
23: em pat.*
24: em ph.*
25: em pl.*
26: em ro.*
27: em scs.*
28: em un.*
29: em vi.*
30: em hcg_hum.*
31: em hcg_inv.*
32: em hcg_other.*
33: em hcg_mus.*
34: em hcg_pln.*
35: em hcg_rnd.*
36: em hcg_mam.*
37: em hcg_vrt.*
38: em sy.*
39: em hcg_hum.*
40: em hcg_mus.*
41: em hcg_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2418.6	100.0	2419	6	AR199602
2	2418.6	100.0	2419	6	AX366860
3	1826.6	75.5	2741	6	AX366876
4	1629	67.3	2475	6	AR199605
5	1629	67.3	2475	6	AX366868
6	1196.6	49.5	2151	6	E01789
7	1196.6	49.5	2151	6	AR205032
8	1196.6	49.5	2151	6	AR282639
9	1196.6	49.5	2151	6	AX082544
10	1196.6	49.5	2151	6	AX370659
11	1196.6	49.5	2315	8	RTPALGNA
12	1196.6	49.5	2331	6	E01783
13	1196.6	49.5	2331	6	E01784
14	1196.6	49.5	2331	6	E01785
15	1196.6	49.5	2439	6	AR199605
16	1196.6	49.5	2439	6	AX366866
17	1195	49.4	2151	6	E01543
18	1195	49.4	2151	6	AR205031
19	1195	49.4	2151	6	AR282638
20	1195	49.4	2151	6	AX082542
21	1195	49.4	2151	6	AX370657
22	1188.8	49.1	2061	6	AR282658
23	1188.8	49.1	2061	6	AX370681
24	1136.6	47.0	2787	6	AR199604
25	1136.6	47.0	2787	6	AX366864
26	802.4	33.2	2986	8	RHPAL2
27	802.4	33.2	3340	8	RHPAL2
28	747.2	30.9	3586	8	RHPAL2
29	402	16.6	2124	6	AX653389
30	402	16.6	2124	6	AX660330
31	402	16.6	2510	8	AK100346
32	401.4	16.6	2015	8	AX224546
33	399.8	16.5	2488	6	MZEPAL
34	399.8	16.5	2126	6	AX660228
35	393.8	16.3	139226	8	OSUNO274
36	393.8	16.3	139488	8	OSUNO262
37	388.8	16.1	170186	2	AC135419
38	388.8	16.1	177374	8	AC135429
39	385.8	15.9	2504	8	AK102817
40	384.2	15.9	2211	6	AX660248
41	384.2	15.9	2280	6	AX658420
42	381.4	15.8	2282	6	AX660254
43	379.8	15.7	2594	8	AK120127
44	379.4	15.7	2157	6	AX654419
45	377.8	15.6	2108	6	AX660252

ALIGNMENTS

RESULT 1
AR199602
LOCUS AR199602 2419 bp DNA
DEFINITION Sequence 12 from patent US 6355468.
ACCESSION AR199602
VERSION AR199602.1 GI:20249676
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2419)
Yoshida, R.X. and Koester, A.B.
Phenylalanine ammonia lyase polypeptide and polynucleotide
sequences and methods of obtaining and using same
JOURNAL Patent: US 6355468-A 12 12-MAR-2002;

FEATURES
source

Location/Qualifiers
1..2419
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 2418.6; DB 6; Length 2419;
Best Local Similarity 100.0%; Pred. No. 1e-284;
Matches 2418; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCCTGCTCACTCAACCCGCTACTAGACTCCGCGATGGCCCTTCCTTGAATCCGCTC 60
Db 1 CTCCTGCTCACTCAACCCGCTACTAGACTCCGCGATGGCCCTTCCTTGAATCCGCTC 60
QY 61 GCCACACGCTCGCCCAAGGCTTTACCAACGGCTCGACGGCTCCGACCAATCGGCT 120
Db 61 GCCACACGCTCGCCCAAGGCTTTACCAACGGCTCGACGGCTCCGACCAATCGGCT 120
QY 121 GCGGGCCCACTTCGAGCTTCGCGCGACGCCGCGCTCGATGCGCACGCCGCGACAG 180
Db 121 GCGGGCCCACTTCGAGCTTCGCGCGACGCCGCGCTCGATGCGCACGCCGCGACAG 180
QY 181 TCGCAGCTCGAATGCTGCGAGGCTCTCAAGCAAGCCCAAGCGAGCTGCTGAGCTC 240
Db 181 TCGCAGCTCGAATGCTGCGAGGCTCTCTCAAGCAAGCCCAAGCGAGCTGCTGAGCTC 240
QY 241 AGCGGATACAGCTTACCGTCCGTGACGTTGTCGCGCGCGCCGCAAGGCGCAGGCTC 300
Db 241 AGCGGATACAGCTTACCGTCCGTGACGTTGTCGCGCGCGCCGCAAGGCGCAGGCTC 300
QY 301 CGGCTCGAAGACGAGAGATCGCGGACGCGCTCGACCAAGACGCTTCCTCAAG 360
Db 301 CGGCTCGAAGACGAGAGATCGCGGACGCGCTCGACCAAGACGCTTCCTCAAG 360
QY 361 GCCCAGCTTCAGACTCGGCTTCAAGGAGTCAACAAGGTTTGGTGGCTCGCGCACG 420
Db 361 GCCCAGCTTCAGACTCGGCTTCAAGGAGTCAACAAGGTTTGGTGGCTCGCGCACG 420
QY 421 AAGACTGAGAGATGAGTCAAGCTTCGAAAGGCGCTCATCGACACAGCTTCGGGGTG 480
Db 421 AAGACTGAGAGATGAGTCAAGCTTCGAAAGGCGCTCATCGACACAGCTTCGGGGTG 480
QY 481 AGCGGACGCTCGGCTTCGCTTCAAGGCTCGAGCGGAGCTCGAGAACAGCTTCGCTC 540
Db 481 AGCGGACGCTCGGCTTCGCTTCAAGGCTCGAGCGGAGCTCGAGAACAGCTTCGCTC 540
QY 541 GAGGTGTCGCGGCGCATGATGCTGCGCTCAACTGCTCAAGCTGAGCTCGGCTC 600
Db 541 GAGGTGTCGCGGCGCATGATGCTGCGCTCAACTGCTCAAGCTGAGCTCGGCTC 600
QY 601 GTCGCGCTGCTGCTTGAAGGCGCTCAACTTCTTGAACCAAGCTCAAGGCTC 660
Db 601 GTCGCGCTGCTGCTTGAAGGCGCTCAACTTCTTGAACCAAGCTCAAGGCTC 660
QY 661 GTCCCTTCGCGGCTTCATCTCGGCGTGGGCGACCTCGAGCTTCCTGATCATGCGC 720
Db 661 GTCCCTTCGCGGCTTCATCTCGGCGTGGGCGACCTCGAGCTTCCTGATCATGCGC 720
QY 721 GCGCGCATCACCGGTCACCGGAGTCAAGTTTGAAGTTTGAACGAGGAAACCGAAG 780
Db 721 GCGCGCATCACCGGTCACCGGAGTCAAGTTTGAAGTTTGAACGAGGAAACCGAAG 780
QY 781 ATCATGTTTGGCGGAGGCTATGCTCTTTGGTCTCGAGGAGTGTCTCGAGCCG 840
Db 781 ATCATGTTTGGCGGAGGCTATGCTCTTTGGTCTCGAGGAGTGTCTCGAGCCG 840
QY 841 AAGGAGGCTTCGCTTCCTGTAACGAAAGCGCTTCGCGCTCGATGAGGAGCCCTCAAT 900
Db 841 AAGGAGGCTTCGCTTCCTGTAACGAAAGCGCTTCGCGCTCGATGAGGAGCCCTCAAT 900
QY 901 CTGACAGCTCGACATGCTCTCGCTCTCTCGAGGCTTGAAGGCTCTCAAGGAG 960
Db 901 CTGACAGCTCGACATGCTCTCGCTCTCTCGAGGCTTGAAGGCTCTCAAGGAG 960

QY 961 GCATGCTCGGCGACGAGGCTCTTGGCCCGTTATCCAGAGCTTGCGCCCGCAC 1020
Db 961 GCATGCTCGGCGACGAGGCTCTTGGCCCGTTATCCAGAGCTTGCGCCCGCAC 1020
QY 1021 CCCGCGCAGTCTGAGGTCGCGCGCATTCGACGCTCTTCCGAGCTGTGCTTGC 1080
Db 1021 CCCGCGCAGTCTGAGGTCGCGCGCATTCGACGCTCTTCCGAGCTGTGCTTGC 1080
QY 1081 GTTGAAGCAGAGGAGGATCAAGTCAAGAGAGGAGGCTTCTTGCCAGAGCGCG 1140
Db 1081 GTTGAAGCAGAGGAGGATCAAGTCAAGAGAGGAGGCTTCTTGCCAGAGCGCG 1140
QY 1141 TACCGCTCGCAGCTCGCTCAATTCCTGCGCGCTGTGTGAGACATGATGACGCG 1200
Db 1141 TACCGCTCGCAGCTCGCTCAATTCCTGCGCGCTGTGTGAGACATGATGACGCG 1200
QY 1201 TACTGACTCTTGTGTCGAGAACACACAGACGACCAACCCGCTCTCGAGCTGAG 1260
Db 1201 TACTGACTCTTGTGTCGAGAACACACAGACGACCAACCCGCTCTCGAGCTGAG 1260
QY 1261 AACAGCAGACCGCGCAGCGGCACTTCAGAGCTGAGCTTCATTTGATGAG 1320
Db 1261 AACAGCAGACCGCGCAGCGGCACTTCAGAGCTGAGCTTCATTTGATGAG 1320
QY 1321 AAGACAGGCTCGCACTGCGCTTCATGCGCAAGCTTCACTGACGAGTGCACCGAGTTG 1380
Db 1321 AAGACAGGCTCGCACTGCGCTTCATGCGCAAGCTTCACTGACGAGTGCACCGAGTTG 1380
QY 1381 CTCAAGCTGTCATGAACCGCGGCTGCTTGTGCTGCTGCTGCGAGGACCGCTGCTC 1440
Db 1381 CTCAAGCTGTCATGAACCGCGGCTGCTTGTGCTGCTGCTGCGAGGACCGCTGCTC 1440
QY 1441 AACTATACCGCAAGGCTTGAACATTCAATGCTGCTTACGCTTCGAGCTCGACAC 1500
Db 1441 AACTATACCGCAAGGCTTGAACATTCAATGCTGCTTACGCTTCGAGCTCGACAC 1500
QY 1501 CTTGCGAACCCGCTGATTAATCTTGTGTCAGCCCGCAAGATGGGCAACAGGCTCGAAC 1560
Db 1501 CTTGCGAACCCGCTGATTAATCTTGTGTCAGCCCGCAAGATGGGCAACAGGCTCGAAC 1560
QY 1561 TCGCTGCTCTCATCTTCGCGCGCCGACACTGCGAGGACCAAGAGCTCTTCTCTCT 1620
Db 1561 TCGCTGCTCTCATCTTCGCGCGCCGACACTGCGAGGACCAAGAGCTCTTCTCTCT 1620
QY 1621 CTGCTTCGACCTGTATCTGACAGCTTCAGGCGCTGACCTTCGCGGATGAGGCTCGAC 1680
Db 1621 CTGCTTCGACCTGTATCTGACAGCTTCAGGCGCTGACCTTCGCGGATGAGGCTCGAC 1680
QY 1681 TTCAAGAGAGGTTGACCCGCTTCTCCGAGCTTCTCTCAGAGACCTCGGCACTGGC 1740
Db 1681 TTCAAGAGAGGTTGACCCGCTTCTCCGAGCTTCTCTCAGAGACCTCGGCACTGGC 1740
QY 1741 CTCGACGTCAACGACTTGGCTTCGAGGTCAGAGTCAAGAGGCTCTCAAGAGGCTCGAGAG 1800
Db 1741 CTCGACGTCAACGACTTGGCTTCGAGGTCAGAGTCAAGAGGCTCTCAAGAGGCTCGAGAG 1800
QY 1801 ACAGCAGCTTACGACCTCGAGCGCGCTGGGACAGAGGCTTCTGTAAGCGAGCGGACCC 1860
Db 1801 ACAGCAGCTTACGACCTCGAGCGCGCTGGGACAGAGGCTTCTGTAAGCGAGCGGACCC 1860
QY 1861 GTGTCAGGCTCTCTGCTCTGCTCGCTTCGCAAGTCAAGCTTACTGCTCGCAACGCG 1920
Db 1861 GTGTCAGGCTCTCTGCTCTGCTCGCTTCGCAAGTCAAGCTTACTGCTCGCAACGCG 1920
QY 1921 TGAAGGTTGCTTCGCGCGAGAGGCTTCTGCTCAAGCTGAGGAGTGTGCAACGCTTC 1980
Db 1921 TGAAGGTTGCTTCGCGCGAGAGGCTTCTGCTCAAGCTGAGGAGTGTGCAACGCTTC 1980
QY 1981 TGGCAGACGCGCTCTTTCGAGGCGCGCGCAGCAAGTACTTCTGCGCGGACCGCGCTC 2040
Db 1981 TGGCAGACGCGCTCTTTCGAGGCGCGCGCAGCAAGTACTTCTGCGCGGACCGCGCTC 2040

Yy		2042	CCTGATCGTTCGTCCGGAGAGACTCGGCGTGACAGGCCGCGCGGCGGACGCTGTTGTC	2100
Dd		2041	CTGTACTCGTTCGTCCGGAGAGACTCGGCGTGACAGGCCGCGCGGCGGACGCTGTTGTC	2100
Oy		2101	GCGCGTCACAGCAGAGCATCGTGGAGCAAGTCCTCCGCATCTTACAGAGCCATCAAGAAC	2160
Dd		2101	GCGGTCCAGCAGAGCATCGTGGAGCAAGTCCTCCGCATCTTACAGAGCCATCAAGAAC	2160
Oy		2161	GCGCGCATTAACCATGTCCTTGTCAAATGTCGCGGTAAGGCCCGAGCAAGCTTGCTTA	2220
Dd		2161	GCGCGCATTAACCATGTCCTTGTCAAATGTCGCGGTAAGGCCCGAGCAAGCTTGCTTA	2220
Zy		2221	GACGCCCCGCTCACCCCAGAACCAGCTTTGAAAGTGTCTGTCGAAGAGCAAGCAAGCTT	2280
Dd		2221	GACGCCCCGCTCACCCCAGAACCAGCTTTGAAAGTGTCTGTCGAAGAGCAAGCAAGCTT	2280
Zy		2281	CCTCATACATCATGTGCGCTTACTCTCTCGCGCGTCATCAGCTCTCAATTCTTTCGAT	2340
Dd		2281	CCTCATACATCATGTGCGCTTACTCTCTCGCGCGTCATCAGCTCTCAATTCTTTCGAT	2340
Zy		2341	CCGCGGCTCTCGGTCGTCCAGTACACGGTAGTAGAGCTCGGAATGATGCAAGCTTCG	2400
Dd		2341	CCGCGGCTCTCTCGGTCGTCCAGTACACGGTAGTAGAGCTCGGAATGATGCAAGCTTCG	2400
Zy		2401	AGTTCAAAAAAAAAAAAAA 2419	
Dd		2401	AGTTCAAAAAAAAAAAAAA 2419	
RESULT 2				
AX366860				
LOCUS	AX366860	2419 bp	DNA	linear PART 15-FEB-2002
DEFINITION	Sequence 12 from Patent WO2008402.			
ACCESSION	AX366860			
VERSION	AX366860.1 GI:18698149			
KEYWORDS				
SOURCE	Rhodotorula graminis			
ORGANISM	Rhodotorula graminis Fungi; Basidiomycota; Urediniomycetes; Microbotryumetidae; Sporidibolales; mtosporic Sporidibolales; Rhodotorula.			
REFERENCE				
AUTHORS	1 Yoshida,R.K. and Kootstra,A.B.			
TITLE	Phenylalanine ammonia lyase polypeptide and polymucleotide sequences and methods of obtaining and using same			
JOURNAL	Patent: WO 0208402-A 12 31-JAN-2002;			
	PGBU Services, Inc. (US)			
FEATURES				
source	location/Qualifiers			
	1..2419			
	/organism="Rhodotorula graminis"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:29898"			
	37..2199			
	/note="unnamed protein product"			
	/codon_start=1			
	/protein_id="CAD23828.1"			
	/db_xref="GI:18698150"			
	/translation="MASELSIATLLANGTNGNSHAAPTKSAAGETSAURTPGLDGHGAAHSGQELIVBELSDPTDDVVELISGYSLRPVVQAARKGRVQCNDDEIFARVRKSDPFLKAOLONSVCVTGTFEGGSADRTEDVAESLLOKALIEHQICVTPISVSFSVGSRGENTPLEVRGMATRVNSELTRGSAVAALVLTLENTLNIRHTIPPLRGSIASASGLDSPLSTIAEATITGPDPVKHYVHEBTEKIMAREALISFLGEAVVLGPKEGELGVNGRAVSAMATSLSDSHMLSLSSQALTALTVEMAVQQGSFPFTIHVCRPHPOQVEVANRITTLSSGSFAVEHEEVKVDDEGIILRODRYPLRLSPQLDELDMWMAVSTLSLENNTDNPDLIDVENKOJAHAGNQPASVVISMEKTRLALALIKNFQCTELNLAAANKNEPSCLAEPDSLNYHGKGLDIHIAVASELGHLANVTTFVPAEKMGDAVNATLSLTSISARITAEANDVLSLIASHLYCTTLCAYDLRAEMEDPFKKDDPLLPITLLOOHGTGDYNALALEYKALKINKRLTEOTTVALPERMDADSYATGTVELLSSS9SANVTLLAVANAUKSAEKAISLTREVARNRWOTPSQAPAAHVISPRTVLXSPFREIELGQARRDVAVGVQOETTGNVRSRIEAIKDRINHVLVKELA"			
CDS				
	/note="Other information: Xaa = Val or Ala"			
	493..495			
unsure				

Query Match	100.0%;	Score 2418.6;	DB 6;	Length 2419;
Best Local Similarity	100.0%;	Pred. No. 1e-28;	Mismatches 0;	Indels 0;
Matches 2419;	Conservative 0;			Gaps 0;
QY 1 CTCTGCTCACTTAACCCGCTACCTAGCACTGCGCATGAGCCCTTCTTGAATCTCCTC	60			
Db 1 CTCTGCTCACTTAACCCGCTACCTAGCACTGCGCATGAGCCCTTCTTGAATCTCCTC	60			
QY 61 GCACCAACGCTCCGCCAACGGCTTTTACCAACGGCTCCGACGCTCCGCTCCGACAACTCGGCT	120			
Db 61 GCACCAACGCTCCGCCAACGGCTTTTACCAACGGCTCCGACGCTCCGCTCCGACAACTCGGCT	120			
QY 121 GCGGGGCCCACTTGGGCTTCGCGCGGACGCTCCGGCTCGATGGGCAACGCGCGGACACAG	180			
Db 121 GCGGGGCCCACTTGGGCTTCGCGCGGACGCTCCGGCTCGATGGGCAACGCGCGGACACAG	180			
QY 181 TCGCAGCTCGAGATCGTGCAGAGACTCTCAGCGACCCCAACCGACGATCGTTCGACTC	240			
Db 181 TCGCAGCTCGAGATCGTGCAGAGACTCTCAGCGACCCCAACCGACGATCGTTCGACTC	240			
QY 241 AGCGGGTTCAGCTTACCTTACCTTCGTTACGTTGTGCGCGCGCCCGCAAGGGGGCAGAGGCTC	300			
Db 241 AGCGGGTTCAGCTTACCTTACCTTCGTTACGTTGTGCGCGCGCCCGCAAGGGGGCAGAGGCTC	300			
QY 301 CGCTCCAGAACGACGACGAGATCCGCGCACCGCTCCACAGAGCGTTCGACTTCTTCAG	360			
Db 301 CGCTCCAGAACGACGACGAGATCCGCGCACCGCTCCACAGAGCGTTCGACTTCTTCAG	360			
QY 361 GCCCACTTACGAATCTGGTTCACGAGTCAACCAACGGGTTTCGGTGGCTCGGCGCAACG	420			
Db 361 GCCCACTTACGAATCTGGTTCACGAGTCAACCAACGGGTTTCGGTGGCTCGGCGCAACG	420			
QY 421 AGCATAGAGATGAGTACGAGCTCCAAAGAGCGTTCATACAGACACAGCTTCGCGCGTG	480			
Db 421 AGCATAGAGATGAGTACGAGCTCCAAAGAGCGTTCATACAGACACAGCTTCGCGCGTG	480			
QY 481 ACGCCGACGTCGCTCTGCTCTTACGGCTGGAACGCGGCTTCGAAACACGCTTCGCTC	540			
Db 481 ACGCCGACGTCGCTCTGCTCTTACGGCTGGAACGCGGCTTCGAAACACGCTTCGCTC	540			
QY 541 GAGGTCTGCGCGCGCGCGCATAGTATCCGCGTCACTCGCTCACGCGTGGCACTTCGACC	600			
Db 541 GAGGTCTGCGCGCGCGCGCATAGTATCCGCGTCACTCGCTCACGCGTGGCACTTCGACC	600			
QY 601 GTCCGCTCTGCTCTTGAAGCGCTCACCAACTCTTGAACCAACCCCATCACGCGCATC	660			
Db 601 GTCCGCTCTGCTCTTGAAGCGCTCACCAACTCTTGAACCAACCCCATCACGCGCATC	660			
QY 661 GTCCCTCTCGCGGCTTCATTTGGCGCTCGGAGCACTCAACCCGCTCTGTAATCGCC	720			
Db 661 GTCCCTCTCGCGGCTTCATTTGGCGCTCGGAGCACTCAACCCGCTCTGTAATCGCC	720			
QY 721 GCGCGCATCAACCGGTCACCGGACGTCAGAGTTTACGTTTGCAGAGGGAACGGAAG	780			
Db 721 GCGCGCATCAACCGGTCACCGGACGTCAGAGTTTACGTTTGCAGAGGGAACGGAAG	780			
QY 781 ATCATGTTTGGCGCGGAGCCCATCTCGCTTTTGGTGTTCAGAGGAGTGTCTCGCCCG	840			
Db 781 ATCATGTTTGGCGCGGAGCCCATCTCGCTTTTGGTGTTCAGAGGAGTGTCTCGCCCG	840			
QY 841 AAGAGAGGTCTCGGTCGTGTCAGGAACGCGCGCTTCGCGCTGAGATGGGAGCCCTCAGT	900			
Db 841 AAGAGAGGTCTCGGTCGTGTCAGGAACGCGCGCTTCGCGCTGAGATGGGAGCCCTCAGT	900			
QY 901 CTGCACGACTCGACATGCTCTCCCTCTTCGACGAGCTTGAAGGCTTCACGAGTGAAG	960			
Db 901 CTGCACGACTCGACATGCTCTCCCTCTTCGACGAGCTTGAAGGCTTCACGAGTGAAG	960			
QY 961 GCATGCTGGCCACAGAGGCTCTTGGCGGCTTATTCACACAGCTTCGCGCCGAC	1020			
Db 961 GCATGCTGGCCACAGAGGCTCTTGGCGGCTTATTCACACAGCTTCGCGCCGAC	1020			

Db 61 CAGCGCGCTCCGACCAAGTCGCGCTCGGGCCCCCACTTCGCTTCGCGCGACGCCCGGCG 120
QY 157 CTCGATGCGCAGCGCGCGACCAAGTCGCGCTCGAGATCGTCGAGAGCTTCACAGGAC 216
Db 121 CTCGATGCGCAGCGCGCGACCAAGTCGCGCTCGAGATCGTCGAGAGCTTCACAGGAC 180
QY 217 CCCACCGACGAGCTGTCGAGCTCAGCGGCTACAGCTCACCGTCGCTGACGTTGTGCGC 276
Db 181 CCCACCGACGAGCTGTCGAGCTCAGCGGCTACAGCTCACCGTCGCTGACGTTGTGCGC 240
QY 277 GCGCGCCCGCAAGGGCGCTCGCGCTCCAGAACGACGAGATCCGCGCACGCGTC 336
Db 241 GCGCGCCCGCAAGGGCGCGAGGCTCGCGCTCCAGAACGACGAGATCCGCGCACGCGTC 300
QY 337 GACAAAGAGCTGACCTTCCTCAAGGCCGAGCTTCAGAACTCGCTACGAGAGTACACAC- 395
Db 301 GACAAAGAGCTGACCTTCCTCAAGGCCGAGCTTCAGAACTCGCTACGAGAGTACACACG 360
QY 396 ----- 395
Db 361 GTGCGTCCGACGACGAGCGGAAATCTCGGATGCGCGACGCGTGAACGCTGACACTC 420
QY 396 -----GCGTTTCGTCGCTCGCGCGACGACGAGGACTGAG 429
Db 421 GCTTGAACGCGCTCGCGCGCTTCGAGAGGTTTCGTCGCTCGCGCGACGAGGACTGAG 480
QY 430 GATGAGCTCAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCGCACG 489
Db 481 GATGAGCTCAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCGCACG 540
QY 490 TCCGCTCTGCTCTTACAGCTGCGGAGCGCGCTCCAGAACAGCTTCGCTCGAGAGTGTCT 549
Db 541 TCCGCTCTGCTCTTACAGCTGCGGAGCGCGCTCCAGAACAGCTTCGCTCGAGAGTGTCT 600
QY 550 CCGCGCGCATGATCATCGCGCTCACTCGCTACGCGTACGCGTGCACACTCGCGCGTCTC 609
Db 601 CCGCGCGCATGATCATCGCGCTCACTCGCTACGCGTGCACACTCGCGCGTCTC 680
QY 610 GTGCTCTTGAAGCGCTACCAACTTCTTGAACCAACGCGATACGCCATCGTCCCTTC 669
Db 661 GTGCTCTTGAAGCGCTACCAACTTCTTGAACCAACGCGATACGCCATCGTCCCTTC 720
QY 670 CCGCGCTCATATCGCGCGTCCGCGCGACCTCAAGCGCGCTTCGTAACATCGCGCGCGCATC 729
Db 721 CCGCGCTCATATCGCGCGTCCGCGCGACCTCAAGCGCGCTTCGTAACATCGCGCGCGCATC 780
QY 730 ACCGCTCAACCCGACGCTCAAGTTCACTTTTGAACGAGGAAACGAGAAATCATGTT 789
Db 781 ACCGCTCAACCCGACGCTCAAGTTCACTTTTGAACGAGGAAACGAGAAATCATGTT 840
QY 790 GCGCGGAGCGCATCTCGCTCTTTGCTCGAGGACATC----- 828
Db 841 GCGCGGAGCGCATCTCGCTCTTTGCTCGAGGACATCTCGAGTCTGACT 900
QY 829 -----G 829
Db 901 GCAAGTGAAGTGTTCAGAGATCTCCAGTTTGTGTACTCCCTTTGTTCAATGCAATTCAG 960
QY 830 TCCCTCGCGCCGAAAGAGAGTCTCGCTCTGTGTCAACGAAACGCGCGTCTCGCGCTCGATG 889
Db 961 TCCCTCGCGCCGAAAGAGAGTCTCGCTCTGTGTCAACGAAACGCGCGTCTCGATG 1020
QY 890 CGACCTCATGCTGACGACTGACAGATGCTCTGCTCTTCGCGAGGCTTCAGCGGCTC 949
Db 1021 CGACCTCATGCTGACGACTGACAGATGCTCTGCTCTTCGCGAGGCTTCAGCGGCTC 1080
QY 950 TCACGCTGAAGGCTCATGCTGCGCGACGAGGCTGCTTCGCGCGTTCATCCACGACGCTC 1009
Db 1081 TCACGCTGAAGGCTCATGCTGCGCGACGAGGCTGCTTCGCGCGTTCATCCACGACGCTC 1140
QY 1010 GCGCGCCGACCCCGCGCACGATCGAGGTCGCGCGCAATCCGCTCTTTCGCGCT 1069

Db 1141 GCCGCCCGACCCCGCGAGTTCGAGTTCGCGCGACCAATCCGACGCTCTTTCGCGCT 1200
QY 1070 CGTGGTTTGCGCTTGAGACGAGAGAGGTCAGAGTCAAGTCAAGAGAGGACTTTC 1129
Db 1201 CGTGGTTTGCGCTTGAGACGAGAGAGGTCAGAGTCAAGTCAAGAGAGGACTTTC 1260
QY 1130 GCCAGACCGGTACCCGCTCGACGTCGCT----- 1161
Db 1261 GCCAGACCGGTACCCGCTCGACGTCGCTCAGTTCGCTCCCTCTCTCCCTTC 1320
QY 1162 -----CAGTTCCTCGCGCGCTCG 1180
Db 1321 CTCGCTCCGACCGCGCGCTGAGACTTAAGTTTTCGATTCAGTTCTTCGCGCTCG 1380
QY 1181 TGAAGACATGATGACGCGCTTACTGACTCTTCGCTCGAGAACCAACACGACGACGACA 1240
Db 1381 TGAAGACATGATGACGCGCTTACTGACTCTTCGCTCGAGAACCAACACGACGACGACA 1440
QY 1241 ACCCGCTCTTGAAGTTCGAGAACCAAGACGCGGACGCGCGCAATTCGAGGCTCGG 1300
Db 1441 ACCCGCTCTTGAAGTTCGAGAACCAAGACGCGGACGCGCGCAATTCGAGGCTCGG 1500
QY 1301 CTGCTCATTTTCAGATGAGAGAGC----- 1325
Db 1501 CTGCTCATTTTCAGATGAGAGAGCAGATGCGCTCTCGCTGCTTCGATCCGATCT 1560
QY 1326 -----CAGGCTCGACTCGCTTCATGCGAGACTCACTTCA 1363
Db 1561 TGTGCTGATGTTCTTCTCTCTCGAGGCTCGACTCGCTTCATGCGAGACTCACTTCA 1620
QY 1364 CGGAGTGAACGAGTGTGTCAGGCTGCATGAACGCGCGCTCGCTTCGCTGCTGCTG 1423
Db 1621 CGGAGTGAACGAGTGTGTCAGGCTGCATGAACGCGCGCTCGCTTCGCTGCTGCTG 1680
QY 1424 CCGAGACCGGTGCTCACTATACGCGCAAGGCTTGAACATTCAGATGCTGCTTACG 1483
Db 1681 CCGAGACCGGTGCTCACTATACGCGCAAGGCTTGAACATTCAGATGCTGCTTACG 1740
QY 1484 CTTGCG----- 1489
Db 1741 CTTGAGAGTGAAGCGGTGACGTTCTCGCGCGTCTCGTCCCTTCAGCGACCCAGGC 1800
QY 1490 -----AGCTCGGACACTTTCGACACCGCGCTGACTACTTCGTCAG 1530
Db 1801 TGACTTCTTTCCTCTGTAAGTTCGCGCACTTTCGCAACCGCGCTGACTACTTCGTCAG 1860
QY 1531 CCGCGACGATGAGCAACGAGCGGTCACTCGCTGCTCATCTTCGCGCGCGCACT 1590
Db 1861 CCGCGACGATGAGCAACGAGCGGTCACTCGCTGCTCATCTTCGCGCGCGCACT 1920
QY 1591 GCGGAGCGCAACGAGCGGTCTTCTC----- 1617
Db 1921 GCGGAGCGCAACGAGCGGTCTTCTCCTGTCGCTTCGTCGCAATGAGTCCGACGCAAT 1980
QY 1618 -----CTTCTCGCTTCGACACTGTACTGACGCTTCAG 1650
Db 1981 AGCGACTGACTGCGCGATCCGAGACAGTTCTGCTCTGACCTGTACTGACGCTCAG 2040
QY 1651 GCGCTGACTTCGCGCGATGAGTGACTTCAAGAGAGATTTCGACCGCTTTCGCG 1710
Db 2041 GCGCTGACTTCGCGCGATGAGTGACTTCAAGAGAGATTTCGACCGCTTTCGCG 2100
QY 1711 ACTCTCTCGAGAGACCTGCGGACTGAGCTGACGATCAACGCACTTCGCGTCAAGTCTC 1770
Db 2101 ACTCTCTCGAGAGACCTGCGGACTGAGCTGACGATCAACGCACTTCGCGTCAAGTCTC 2160
QY 1771 AAGAGGCGCTCAACAGCGTCTCGAGACGACGAGTGAACGACTTCGAGCGCGCTCG 1830
Db 2161 AAGAGGCGCTCAACAGCGTCTCGAGACGACGAGTGAACGACTTCGAGCGCGCTCG 2220
QY 1831 CACGACGCTTCTGTAAGGACCGGCACTGCTGAGAGCTTCCTCTGCTGCTGCTCT 1890
Db 2221 CACGACGCTTCTGTAAGGACCGGCACTGCTGAGAGCTTCCTCTGCTGCTGCTCT 2280

QY 1891 GCCAAGTACACCTTACTGCGGTCAAACGGTGGAAAGTTGCTTCGGCCGAGAGGCCATC 1950
DB 2281 GCCAAGTACACCTTACTGCGGTCAAACGGTGGAAAGTTGCTTCGGCCGAGAGGCCATC 2340
QY 1951 TCGCTCAGCGGAGGTGGCAACGGCTTCTGGGAGAGCGCGTCTTGGCAGGCGCGCG 2010
DB 2341 TCGCTCAGCGGAGGTGGCAACGGCTTCTGGGAGAGCGCGTCTTGGCAGGCGCGCG 2400
QY 2011 CACGATACCTCTCGCCGCGCAACGGCTCTGTAAGTCTGTTGCGGAGAGCTCGG 2070
DB 2401 CACGATACCTCTCGCCGCGCAACGGCTCTGTAAGTCTGTTGCGGAGAGCTCGG 2460
QY 2071 GTGCAAGCGCGCGCGCGCAACGTGTTGTTGCGCGTGGCAGAGAGAGATCGGAGAAC 2130
DB 2461 GTGCAAGCGCGCGCGCGCAACGTGTTGTTGCGCGTGGCAGAGAGAGATCGGAGAAC 2520
QY 2131 GTCTCGCATCTACAGAGCGCATCAAGAGCGCGCATCAACAGCTCTGTCAGATG 2190
DB 2521 GTCTCGCATCTACAGAGCGCATCAAGAGCGCGCATCAACAGCTCTGTCAGATG 2580
QY 2191 CTGCGTAAAGCCCGAGCAAGCGCTCGCTTAAGCGCCGCTCACCCCAAGCAGCTTTT 2250
DB 2581 CTGCGTAAAGCCCGAGCAAGCGCTCGCTTAAGCGCCGCTCACCCCAAGCAGCTTTT 2640
QY 2251 CGACGTGTGTGTGTGCAAGAGCGAGCTTCTCTCATACATGTGCGCTTACTCTCG 2310
DB 2641 CGACGTGTGTGTGTGCAAGAGCGAGCTTCTCTCATACATGTGCTTACTCTCG 2700
QY 2311 CGGTCAACAGTCTCTCAGTTCTTTGATCCCGGCTCT 2351
DB 2701 CGGTCAACAGTCTCTCAGTTCTTTGATCCCGGCTCT 2741

RESULT 4
AR199606 2475 bp DNA linear PAT 20-APR-2002
LOCUS AR199606 Sequence 20 from patent US 6355468.
DEFINITION AR199606
ACCESSION AR199606
VERSION AR199606.1 GI:20249680
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2475)
AUTHORS Yoshida,R.K. and Kooststra,A.B.
TITLE Phenylalanine ammonia lyase polypeptide and polynucleotide
JOURNAL Patent: US 6355468-A 20 12-MAR-2002;
FEATURES
source 1..2475
/organism="Unknown"
/mol_type="unasigned DNA"

ORIGIN
Query Match 67.3%; Score 1629; DB 6; Length 2475;
Best Local Similarity 78.0%; Pred. No. 7,5e-189;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

QY 37 ATGGCCCTCTCTTGAATCGCTGCGCAACAGCTGCGCAACGGCTTACCAAGCGCTCG 96
DB 1 ATGGCCCTCTCTTGAATCGCTGCGCAACAGCTGCGCAACGGCTTACCAAGCGCTCG 60
QY 97 CAGCGCGCTTCGACCAAGTGGCTGGCGGCCCACTTGGCTCTCGCGCAGCAGC--CCG 154
DB 61 CAGCGCGCTTCGACCAAGTGGCTGGCGGCCCACTTGGCTCTCGCGCAGCAGCAGC 120
QY 155 GCGTCGATGGCCAGCGCGCGCAACAGTGGCGCTGCGATGTGTCAGAGAGCTCTCAGCG 214
DB 121 GCGTCGCTCTCCGCAACCCAGNNAGAGCGAGCTCGACATGTMGAGNAGATCTCGCG 180
QY 215 ACCCGACGAC---GACGTGTGAGCTCAGCGGAGTACAGCTCAGCGTCCGTGACGTTG 271

DB 181 ACCCGACGNNACGNNACGNNNTGCAACTCGACGGGTACACCTTCACCTTCGNGACTCG 240
QY 272 TCGGCGCGCCCGGAGAGGGCGCAGGGTTCGCGTC-CAGAACGACGACGAGATCCGCGA 330
DB 241 TCGGCGCGCCCGGAGAGGGCGCAGGGTTCGCGTC-CAGAACGACGAGATCCGCGA 300
QY 331 CGGTCGACAGAGCGTCACTTCTCTCAAGGCCAGTTTAAAGTCTGCTTACGAGATC 390
DB 301 AAGATGACAAAGAGTGAAGTCTCTCGNNCAGCTCAACACAGNGTCTACGAGATC 360
QY 391 ACCAGGGTTCGAGTGGCTCGGCCGACACAGAGACTGAGAGATGACGCTCAGAG 450
DB 361 ACGATGATTCGCGGCTCGGCCGACACCGGACATGAGATGCAATCTTCGCTCCAGAG 420
QY 451 GCGCTCATGACACCAAGCTTCGCGGCTGAGCGCGACGCTCGCTCTGCTTCAAGCTC 510
DB 421 GCGCTCATGACACCAAGCTTCGCGGCTGAGCGCGACGCTCGCTCTGCTTCAAGCTC 480
QY 511 GAGCGCGGCTTCGAGAACAGGCTTCGCGTGAAGTGTGCGGCGGCGCATGTCATCGGC 570
DB 481 GAGCGCGGCTTCGAGAACAGGCTTCGCGTGAAGTGTGCGGCGGCGCATGTCATCGGC 540
QY 571 GTCAACTGCTCAGCGGTGCGCACTCGGCCGCTCGGCTCGTCTCTTGAAGCGCTCAC 630
DB 541 GTCAACTGCTCAGCGGTGCGCACTCGGCCGCTCGGCTCGTCTCTTGAAGCGCTCAC 600
QY 631 AACTTCTTGAACACCGCATACAGCCGATGTCCTCCCTCGCGGCTTCATTCGCGCTG 690
DB 601 AACTTCTTGAACACCGCATACAGCCGATGTCCTCCCTCGCGGCTTCATTCGCGCTG 660
QY 691 GCGGACCTCAGCCGCTCTGTAACATGCGCGGCGGACATCAGCGTCAACCGACGTAAG 750
DB 661 GCGGACCTCAGCCGCTCTGTAACATGCGCGGCGGACATCAGCGTCAACCGACGTAAG 720
QY 751 GTTACGTTTTCACAGAGGAGAACCGAGAGATATGTTTTCGCGCGAGGCTTCGCTC 810
DB 721 GTTACGTTTTCACAGAGGAGAACCGAGAGATATGTTTTCGCGCGAGGCTTCGCTC 780
QY 811 TTGAGTCTCAGAGGAGTGTCTCGCGCGGAGAGAGGCTGTGTCACAGGAGAG 870
DB 781 TTGAGTCTCAGAGGAGTGTCTCGCGCGGAGAGAGGCTGTGTCACAGGAGAG 840
QY 871 GCGGCTTCGCGCTCGATGAGGAGCGACCTCACTGTCACAGTCTCGCATCTGCTCTC 930
DB 841 GCGGCTTCGCGCTCGATGAGGAGCGACCTCACTGTCACAGTCTCGCATCTGCTCTC 900
QY 931 TCGCAGGCTTGAAGGCTTCAACGTTGAGAGCCATGTTGGCCAGAGAGGCTTCGCT 990
DB 901 TCGCAGGCTTGAAGGCTTCAACGTTGAGAGCCATGTTGGCCAGAGAGGCTTCGCT 960
QY 991 CCGTTCATTCACAGAGCTCTCGCGCGCACCGCGCAGGCTGAGGTCGCGGCAACATC 1050
DB 961 CCGTTCATTCACAGAGCTCTCGCGCGCACCGCGCAGGCTGAGGTCGCGGCAACATC 1020
QY 1051 CGCAGGCTCTTCGCGCTGCTGCTGTTGGCTTGAAGACGAGAGAGGTCAGAGTCAAG 1110
DB 1021 CGCAGGCTCTTCGCGCTGCTGCTGTTGGCTTGAAGACGAGAGAGGTCAGAGTCAAG 1080
QY 1111 GACGACGAGGAGCATTTCTTGCAGAGACCGGCTACCGGCTCGACGTCGCTCAGTTCCTC 1170
DB 1081 GACGACGAGGAGCATTTCTTGCAGAGACCGGCTACCGGCTCGACGTCGCTCAGTTCCTC 1140
QY 1171 GCGCGCTCTGTAAGGATGATGACAGGCTTCTGCTGCTGAGAAC--AAC 1227
DB 1141 GCGCGCTCTGTAAGGATGATGATGACAGGCTTCTGCTGCTGAGAAC--AAC 1200
QY 1228 ACGACGACGACGACACCGGCTCTCGACGTCGAGAACACACGACGCGGAGCGGCGAAC 1287
DB 1201 TCGACGACGACGACACCGGCTCTCGACGTCGAGAACACACGACGCGGAGCGGCGAAC 1260
QY 1288 TTCCAGGCGTGTGCTGCTGATTTTCATGAGAGAACGAGGCTCGCATCTCGCTCATC 1347
DB 1261 TTCCAGGCGTGTGCTGCTGATTTTCATGAGAGAACGAGGCTCGCATCTCGCTCATC 1320

QY 631 AACTTTGAACACGCGATCAAGCCCAATCGTCCCTCGGAGGCTCATCTCGGCGTGC 630
 DB 601 AACTTCTCAACAGGCGATCAACCCCAATCGTCCCTCGGAGGCGATCTCGGCGTGC 660
 QY 691 GCGGACCTGAGCCGCTCTGTACATCGCGCGCCATCAACCGTCAACCCGAGCTCAG 750
 DB 661 GCGGACCTCTCCCTCTCTCATATCGCGCGCCATCAACCGTCAACCCGAGGAG 720
 QY 751 GTTACGTTTGAAGAGGAGAACCGAAGATCATGTTGGCGCGAGGCGCATCTCGCTC 810
 DB 721 GTTACGTTTGAAGAGGAGAACCGAAGATCATGTTGGCGCGAGGCGCATCTCGCTC 780
 QY 811 TTTGCTCTGAGAGGAGTCTGCTCGGCGCGAGAGGAGTCTGCTCTGCTCAACGAGAG 870
 DB 781 TTTGCTCTGAGAGGAGTCTGCTCGGCGCGAGAGGAGTCTGCTCTGCTCAACGAGAG 840
 QY 871 GCGGCTCTGCTCTGATGAGGAGCCTCAAGTCTGAGAGCTGAGCATGCTCTGCTCTC 930
 DB 841 GCGGCTCTGCTCTGATGAGGAGCCTCAAGTCTGAGAGCTGAGCATGCTCTGCTCTC 900
 QY 931 TCGAGAGCCTTGAAGGAGTCTCAAGTGAAGGAGTGAAGTGAAGGAGGAGTGAAGG 990
 DB 901 TCGAGAGCCTTGAAGGAGTCTCAAGTGAAGGAGTGAAGTGAAGGAGGAGTGAAGG 960
 QY 991 CGGTTCACTCAAGAGTCTGCGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1050
 DB 961 CGGTTCACTCAAGAGTCTGCGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
 QY 1051 CGGAGGCTCTTCCGCGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1110
 DB 1021 CGGAGGCTCTTCCGCGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
 QY 1111 GACGAGAGGAGGAGTCTTCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1170
 DB 1081 GACGAGAGGAGGAGTCTTCCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
 QY 1171 GCGCGGCTCTGAGAGGAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1227
 DB 1141 GCGCGGCTCTGAGAGGAGTGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
 QY 1228 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1287
 DB 1201 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
 QY 1288 TTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1347
 DB 1261 TTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
 QY 1348 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1407
 DB 1321 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
 QY 1408 CCTTGTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1467
 DB 1381 CCTTGTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
 QY 1468 CAGCATGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1527
 DB 1441 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
 QY 1528 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1587
 DB 1501 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
 QY 1588 ACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1647
 DB 1561 ACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
 QY 1648 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1707
 DB 1621 CAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
 QY 1708 CCGACTCTCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1755

DB 1681 NNNNGCTCTCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 QY 1756 CTTCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1815
 DB 1741 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
 QY 1816 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1875
 DB 1801 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
 QY 1876 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1933
 DB 1861 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
 QY 1934 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1993
 DB 1921 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
 QY 1994 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2053
 DB 1981 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
 QY 2054 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2113
 DB 2041 CTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100
 QY 2114 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2173
 DB 2101 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
 QY 2174 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2232
 DB 2161 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
 QY 2233 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2292
 DB 2221 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2280
 QY 2293 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2352
 DB 2281 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2340
 QY 2353 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2411
 DB 2341 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
 QY 2412 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2479
 DB 2401 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2460

RESULT 6
 E01759
 LOCUS DNA encoding L-phenylalanine ammonia-lyase. 2151 bp RNA linear PAT 29-SEP-1997
 DEFINITION
 ACCESSION E01759
 VERSION E01759.1 GI:2170012
 KEYWORDS JP 1988291583-A/1.
 SOURCE Rhodospiridium toruloides
 ORGANISM Rhodospiridium toruloides
 Microbiota; Fungi; Basidiomycota; Urediniomycetes; Rhodospiridium.
 1 (bases 1 to 2151)
 REFERENCES
 Fukunaga, N., Yoshino, S., Yamamoto, K., Nakajima, Y., Iwata, T.,
 Matsumoto, N., Suzuki, M., and Wakiguchi, N.
 AMINO ACID SEQUENCE OF L-PHENYLALANINE AMMONIA LYASE, ITS
 STRUCTURAL GENE, NOVEL BASE SEQUENCE CONTAINING SAME, TRANSFORMANT
 PRODUCED THEREWITH AND PRODUCTION OF L-PHENYLALANINE USING SAME
 Patent: JP 1988291583-A 1 29-NOV-1988;
 MITSUBI TOYOTSU CHEM INC
 OS Rhodospiridium toruloides
 COMMENT
 PN JP 1988291583-A/1

Db 1725 GAAGGTGAACAAGAGCTGCGCAAGCGCTCGAGCAGAACCACTGTGACGACTCTGCC 1784
QY 1824 GCGCTGACAGAGCGCTTCTGTAGCGACCGGACCGGTGCTGAGCTCTCTGCTC 1883
Db 1785 GCGCTGACAGAGCGCTTCTGTAGCGACCGGACCGGTGCTGAGCTCTCTGCTC 1844
QY 1884 GCGCTTGCACAGCTGACCGCTTACTGCGGTCAACCGGTGAGAGTTCCTGCGCGAGAA 1943
Db 1845 GTC-----GCTCTGCTGCGCGCGGTCAAGCGCTGAGAGGTGCGCGCGAGATC 1895
QY 1944 GCGCATCTGCTGACGCGCGAGAGTGCACACCGCTTCTGAGCAGCGCGCTTCTGAGAGC 2003
Db 1896 GCGCATCTGCTGACCGCGAGAGTGCACACCGCTTCTGAGCAGCGCGCTTCTGAGAGC 1955
QY 2004 GCGCGGACAGCATCTCTGCGCGGACCGGACCGGCTGCTGACTGCTTCTGCGCGAGGA 2063
Db 1956 GCGCGGCTCTGCTGCTGCTGCTGCGCGGACCGGCTGCTGACTGCTTCTGCGCGAGGA 2015
QY 2064 GCTCGGCTGACAGCGCGCGCGCGGACCGGTGCTGCTGCTGCGCGAGAGAGATCGG 2123
Db 2016 GCTTGGCGTCAAGCGCGCGCGCGGAGAGCTTCTGCGCGAGAGAGAGTGAAGATCGG 2075
QY 2124 GAGCAAGCTCTGCGCGCTTCTGAGAGCGCATCAAGAGAGCGCGCATCAAGCGCTCTCT 2183
Db 2076 CTCGAACTGCTCAAGATCTTACGAGGCGCATCAAGTCCGCGAGATCAACAGCTCTCT 2135
QY 2184 CAAGATGCTCGCGTA 2198
Db 2136 CAAGATGCTCGCTTA 2150

RESULT 7
AR205032 2151 bp DNA linear PAR 20-JUN-2002
LOCUS AR205032
DEFINITION Sequence 9 from patent US 6368837.
ACCESSION AR205032
VERSION AR205032.1 GI:21502510
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2151)
AUTHORS Gatenby/A.A., Sarsisiani,S., Tang,X.-S., Qi,W.Wei. and Vannelli,I.T.
TITLE Bioproduction of para-hydroxycinnamic acid
JOURNAL Patent: US 6368837-A 9-09-Apr-2002;
FEATURES
Location/Qualifiers
1..2151
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-116;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 39 GCGCGCTGCAAGAGTGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 158
Db 45 CCGTCCGATCCGCAAGAGCGCTGCTCAATGCGCGCTGCAACCACTGCGAGTGCAGAGCTC 104
QY 159 CGATGCCACCG 218
Db 105 GCACCTGCGCGCAACCG 164
QY 219 CACCGAGAGCTGCTGAGCTCAGCGGCTACAGCTCAAGCTCGCGCGCGCGCGCGCGCGCG 278
Db 165 GACGATCTGAGCGCTGAGCTGAGCGGCTACCTGCTCAACTCGAGAGCGTCTCTCGGC 224
QY 279 GCGCGCAAGGCG 338
Db 225 GCGCGAAGGCG 284
QY 339 CAAGAGCTGCACTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398

Db 285 CAATTCGCTCGAGTCTTTCGCTCGCAACTCTCCATGAGGCTGACGCGTCAAGAGCTCG 344
QY 399 TTTGGTGGCTCG 458
Db 345 ATTGGCGGATCCGCAACACCGCGACCGAGAGCGCGCATCTCGCTCAAGAGCTCTCT 404
QY 459 CGAGCAGAGCTTGCAGCG 518
Db 405 CGAGCAGAGCTTGCAGCG 464
QY 519 CTTGAGAAACGCTTTCGCTGAGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 578
Db 465 TCTGAGAACTGCTTTCGCTGAGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524
QY 579 GCTCAGCGGTGGCG 638
Db 525 CTTAGCCCGGCG 584
QY 639 GAACCCAGCATCG 698
Db 585 CAACCCAGCATCG 644
QY 699 CAGCGCGCTCTGCTACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 758
Db 645 CT 704
QY 759 TTTCACAGAGGAAACCGAAGATCAATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCG 818
Db 705 CGTCACAGAGGAAACCGAAGATCAATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 764
QY 819 CGAGCAGTCTGCTCTGCG 878
Db 765 GAGCGCGCTGCTCTGCG 824
QY 879 CGCTTCATGCGCAACCTCAGTCTGACAGCTGCAATGCTCTGCTCTCTCTCTCTCTCT 938
Db 825 AGCATGATGCGCAACCTCAGTCTGACAGCTGCAATGCTCTGCTCTCTCTCTCTCTCT 884
QY 939 CTTGACGCTCTGACGCTGAGAGCGCATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 998
Db 885 GCTCAGCGCATACGCTCGAAGCGATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 944
QY 999 CCAAGAGCTGCG 1058
Db 945 TCAAGAGCTACCG 1004
QY 1059 CTTTCCGCTGCTGCTGCTTTCGCTGAGCAGAGAGAGAGTCAAGTCAAGAGAGAG 1118
Db 1005 CTTGAGAGAGAGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
QY 1119 GGGATCTCTGCGCAGAGCGCGCTACCGCGCTCGCGCAGTGTGCTCTGCTGCTGCTGCT 1178
Db 1065 GGGATCTCTGCGCAGAGCGCGCTACCGCGCTCGCGCAGTGTGCTCTGCTGCTGCTGCT 1124
QY 1179 CGTGAAGAGCATATGACCGCTACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db 1125 GCTGAGGAGCATATGACCGCTACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
QY 1236 CGACACCGCTCTCTGAGTGCAGAAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCG 1295
Db 1185 CGACACCGCTCTCTGAGTGCAGAAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCG 1244
QY 1296 GTCGCGCTGCTGATTCGATGAGAGAGCGAGGCTCGCACTGCGCGCTCATGCGCAAGCT 1355
Db 1245 TCGCGCTGCTGAGTGCAGAAACAGAGACCGCGCTCGCGCGCGCGCGCGCGCGCG 1304
QY 1356 CAATTCAGCAGTGCAGCGAGTGTCTAAGCTGCTCCATGAACCGCGCGCTGCTGCTG 1415
Db 1305 CAATTCAGCAGTGCAGCGAGTGTCTAAGCTGCTCCATGAACCGCGCGCTGCTGCTG 1364
QY 1416 CTTGCGTCCGAGAGCGCGTCTGCTCAATGACGCGCAAGGCTTGGCATTTCACTGCG 1475
Db 1365 CTTGCGTCCGAGAGCGCGTCTGCTCTTCACTGCAAGGCTTGGCATTTCACTGCG 1424

1476 TGCTTAGCGCTTGGAGCTCGGCGACCTTGCCAAACCGGCTACATCCTTGCTCCAGGCGCC 1535
1425 GCGGAGACCTCTGGAGGTGGGAGACCTTCGCCAACCTGTGACAGACGATGTCCAGCGCGC 1484
1536 AGAGATGGGACCAACGAGCGCGTCAACTGCTGCTGTATCTCCGCGCGCGGACATGCGCA 1595
1485 TGAATGAGCGAACCAGCGGCTCAACTGCTTGCGCTATCTCGGCTGTGTGACAGACCA 1544
1596 GCGCAACGAGCTCCTTTCT 1655
1545 GTCCAAACGAGCTCTTCT 1604
1656 GCACTCCGCGCGGATGAGAGCTCGACTTCAAGAAAGATTTGACATCCCGCTTCTCCGACTCT 1715
1605 CGACTTGGCGCGGAGCCGAGTTGAGTTCAAGAACGATTCGCGCCAGCCATCGTCTCGCT 1664
1716 CCTCCAGAGCACTCTGAGCACTGAGCTCTGACGTCA-----ACGCACTTGCGCT 1763
1665 CATGACACGAGCACTTTGGCTCCGCCATGACCGGCTCGAACCTTGCGGAGCGAGCTCGTCA 1724
1764 CGAGGTCAAGAGGCGCTCAACAGCGCTCTCGAGAGACAGACGATGACGACTCGAGCC 1823
1725 GAAAGTGAACAAGACGCTCGCAAGCGCTTCGAGCAGACCACTCGTACGACTCTGTCCTC 1784
1824 GCGGTGGAGACGCGCTTCTCTGTAACGAGACCGGACCGCTGTCGAGCTCTCTCTCTCTCTCT 1883
1785 GCGCTGGACACAGCGCTTCT 1844
1884 GCGCTCTGCAACGCTCAACCTTACTGCGCGTCAACGCGTGAAGGTTGCTCTCGGCGAGAA 1943
1845 GTC-----GCTCTGCTCTGCGCGCGTCAACGCTGGAAGGTGCGCGCGCGCGAGTCTC 1895
1944 GGCATCTCTGCTCAAGCGCGAGGTGGCAACCGTTTGGCAAGCGCGCTTCTGCAAGGC 2003
1896 GGCATCTCTGCTCAAGCGCGAGGTGGCAACCGTTTGGCAAGCGCGCTTCTGCAAGGC 1955
2004 GCGCGCGACGACATCTCTGCGCGCGACGCGGCTCTGATCTGTTCTGTCGCGAGGA 2063
1956 GCGCGCGCTCTCTGATCTCTCTGCGCGGACATCAATCTCTAGCTCTCTCTCTCTCTCTCTCT 2015
2064 GCTCGGCGTGAAGGCGCGCGCGGAGCGTCTTGTGTGCGCGTGAAGAGACGATCGG 2123
2016 GCTTGGCTCAAGGCGCGCGCGGAGCGTCTTGTGTGCGCGTGAAGAGGTTGACGATCGG 2075
2124 GAGCAAGCTCTGCGGACGCTACGAGGCGCATCAAGAGAGCGCGGATCAACGATCGCTCGT 2183
2076 CTCGAGCTCTTCCAGATCTTACGAGGCGCATCAAGTGGGCGAGATCAACGATCTCTCTCT 2135
2184 CAAGATGCTCGCGTA 2198
2136 CAAGATGCTCGCTTA 2150

RESULT 8
AR282639 2151 bp DNA linear PAT 10-APR-2003
LOCUS AR282639
DEFINITION Sequence 9 from patent US 6521748.
ACCESSION AR282639
VERSION AR282639.1 GI:25719238
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 2151)
Tang,X.-S.
TITLE Polynucleotide encoding a mutant Rhodotorula glutinis tyrosine
ammonia lyase polypeptide
JOURNAL Patent: US 6521748-A 9 18-FEB-2003;
FEATURES Location/Qualifiers
1..2151
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
99 GCGCGCTCCGACCAAGTCGCGCTCGGCGCCCACTTGGGCTCTCCGCGGACGCGCGGCT 158
45 GGTGATCTCGCAAGAGAGGCTGTCAATGGCGCTCGACCAACTGTGCAAGTGCAGAGCTTC 104
159 CGATGGCCAGCGCGCGCACGATCGAGCTCGAATGTGACAGAGCTCTCTCGACGACCC 218
105 GCACCTGCCCAACCAACGAGTCAAGAGGTGACATGTGAGAAAGATGTCTGCGCGCC 164
219 CACCGACAGCTGTGAGCTCAAGCGGTGACCTTCAACCTCCGTCGAGAGTTTGTGCGCC 278
165 GACCGACTCGACGCTGGAACCTGACGCGCTACTGCTCAACTCTCGAGAGCTGTCTCGGC 224
279 GCGCGGCAAGGAGGAGCGAGGCTCGCGTCCGATCGAAGACGAGACGAGATCCGCGACGCTCGA 338
225 GCGGAGGAGGAGGAGGCTGTCTCGCGTCAAGAGCAGCGAGATCTCGCTCAAGATTGA 284
339 CAAGAGCGTGAATCTTCTCAAGGCCAGCTTCAAGAACTGGTCTACGAGTCAACAGG 398
285 CAATCGGTGAGATTCTTGGCGCTCGCAACTCTCCATGAGCGTCTACGCGGTCAAGCTG 344
399 TTTGGTGGCTCGGCGCGACAGAGACTGAGATGAGTCAAGCTCGCTCAGAAAGCGCTCAT 458
345 ATTGGAGGATCCGACACACCGGACCGAGAGCGCATTCGCTCCAGAAAGCTCTTCTCT 404
459 CGACACAGCTCTGCGCGCTGAGACCGGAGCTGCTCTGCTCTTCAAGGCTCGGAGCGG 518
405 CGAGACACAGCTCTGCGGCTGTCT 464
519 CCTGGAAGACAGCTTCTCGCTCGAGTGTCTCGCGGCGCATAGTATCTCGGCTCAACTC 578
465 TCTGAGAACTCGTCTCTCTCTCGAGTTGTGCGGCGCATCAATCGGCTCAACAG 524
579 GCTCAGCGTGGCCACTCGGCGCGCTCTGCTCTGTTAGAGCGCTCAACAACTTCTT 638
525 CTGACCGCGGCGCACTCGGCTGTCTGCGCTCTGCTCTGAGGCGCTCAACAACTTCTT 584
639 GAACACCGCATCAAGCGCATTCGCTCTCGCGGCTCTCATCTCGGCGTGGGCGACT 698
585 CAACACCGCATCAAGCGCATTCGCTCTCGCGGCTCTCATCTCGGCGTGGGCGACT 644
699 CAGCGCGCTCTGATCAATCGCGCGGCGCATCAAGGATCAACCGGATCAAGGTTACGCT 758
645 CTCTCTCTCTCTCAATGAGAGGCGCATCAAGGATCAACCGGATCAAGGTTACGCT 704
759 TTTGACGAGGGAACGAGAAAGATCATGTTTGGCGCGGAGCCATCTCGCTCTTGTGCT 818
705 GGTCAAGAGGGAAGAGAAAGATCTGTACGCGCGGAGGCGATGCGGCTTTCACCT 764
819 CGAGGCAATGCTCTCTGCGCGCGGAGAGGCTCTGCTGTGTCAAGGAAAGGCGCTCTCT 878
765 GAGCGCGCTCTCTCTGCGCGCGGAGAGGCTCTGCTGTGTCAAGGAAAGGCGCTCTCT 824
879 GCGCTGATGCGGAGCTCTGATGTCAGACTCGCATCTCTGCTCTCTCTCTCTCTCTCTCTCT 938
825 AGCATGATGCGGAGCTCTGATGTCAGACTCGCATCTCTGCTCTCTCTCTCTCTCTCTCTCT 884
939 CTGACGCTCTCTCAAGTGAAGGCGATGTGCGGACGAGGCTCTGCTGCGCGCTTCAT 998
885 GCTACGCGCATGACGATGCAAGGATGTGCGGCGACGCGGCTCTGCTCACCCCTTCCT 944
999 CCAGAGGTGTCGCGCGCGGACCGCGGCGAGGTGAGGTGCGCGGCAACTCGCGAGCT 1058
945 TCAGAGGTGTCGCGCGCTCAACCGGACGAGTGAAGTGTGCGGAAACATCTCGCAAGCT 1004
1059 CTTTCCGCTCTGCTGTTTCCGTTTGAAGACGAGAGAGTCAAGGTCAAGGACGACGAG 1118
1005 CTTGAGGGAAGCGGCTTGTGTCTGTCACATGAGAGGAGTCAAGGTCAAGGACGACGAG 1064

QY 1119 GGGCATTTCTTCCAGAGACCGCTACCGCTCGCGACGTCGCGCTCGAGTTCCTCGGCCGCT 1178
DB 1065 GGGCATTTCTTCCAGAGACCGCTACCGCTCGCGACGTCCTTCTAGTGGCTCGGCCGCT 1124
QY 1179 CGTGAGAGCATGATGACAGCGCTTCTGATCTCTCGCTCGAGAACAA---CGACGAC 1235
DB 1125 CGTACAGCAGCTCATTTACAGCCCGCTCTCAACATCGAGCGCCGACGATGAGAC 1184
QY 1236 CGACACACCGCTCTCGACGTCGAGAACAGACAGACCGCCGACGCGCGCACTTCCAGGC 1295
DB 1185 CGACACCTCTCTCATGAGTCGAGAACAAAGATTGCAACACGCGCGCAATTTCCAGGC 1244
QY 1296 GTGCGCTGTCTGATTTGATGAGAACAGACGATCGCATCGCCCTCAATCGGACGCT 1355
DB 1245 TGCCTGTGTGCGAACACCATGAGAACACTCGCTGTGGGCTGCGCCAGATCGGCAAGCT 1304
QY 1356 CAATTTACAGCAGTGCAGCGAGTTGCTCAACGCTGCGCATGACCGCGCGCTGCTTGTG 1415
DB 1305 CAATTTACAGCAGCTCACGAGATGCTCAACGCGCGCATGACCGCGCGCTTCTCTG 1364
QY 1416 CTTGCGTCCGAGAGACCGCTCGCTCAATATCAAGGAGGCTTGAACAATTCATCGC 1475
DB 1365 CTTGCGGCGCGAAGACCTCTGCTCTTACATGCAAGGCTTCAACATCGCGCTGCTG 1424
QY 1476 TGCCTACGCTTGAAGCTCGGACCTTGCACCTTGCACCTGCTGCTGCTGCTGCTGCTG 1535
DB 1425 GCGCTACACCTCGAGATTGGGACACCTGCGCACTGTGACGATGCTCAGCGGCG 1484
QY 1536 AAGAGTGGGCAACGAGCGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1595
DB 1485 TGAAGTGGGCAACGAGCGGCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1544
QY 1596 GGCACAGCAGTCTTCT 1655
DB 1545 GTCCACAGCAGTCTTCT 1604
QY 1656 GCACTTCGCGCGAGTGCAGTCTCAAGAACAGTTCGAGCCAGCCTGCTGCTGCTGCTG 1715
DB 1605 GCACTTCGCGCGAGTGCAGTCTCAAGAACAGTTCGAGCCAGCCTGCTGCTGCTGCTG 1664
QY 1716 CTTCCAGCAGCCTCGGCACTGGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1763
DB 1665 CATGACAGCAGCTTGGCTCGGCTGCACTGCACTGCACTGCACTGCACTGCACTGCACTG 1724
QY 1764 CGAGTCAAGAAAGCGCTCAACAGCTCTGAGACAGACAGACGATGCACTGCACTGCACTG 1823
DB 1725 GAAGGTGAACAGAGCTGCGCAAGCTGCGAGCAGACGATGCACTGCACTGCACTGCACTG 1784
QY 1824 GGGCTGGACAGCGCTCTCTGTAAGCGAGACGCGTCTGAGTCTCTGCTGCTGCTGCTGCTG 1883
DB 1785 GGGCTGGACAGCGCTCTCTGTAAGCGAGACGCGTCTGAGTCTCTGCTGCTGCTGCTGCTG 1844
QY 1884 GGCCTTGTGCAACGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1943
DB 1845 GTC-----GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1895
QY 1944 GGCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2003
DB 1896 GGCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1955
QY 2004 GCGCGGACGAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2063
DB 1956 GCGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2015
QY 2064 GCTCGGCTGAGCG 2123
DB 2016 GCTTGGCGTCAAGGCG 2075
QY 2124 GAGCAAGCTGCGGCTCTGCAAGGCGATCAAGAGCGCGCGCTCAAGCGCTGCTGCTGCTG 2183
DB 2076 CTGCAAGCTTCTCAAGATCTAGAGGCGATCAAGTCTGCGGCGAGATCAAGACGCTCTCTCT 2135

QY 2184 CAAGATGCTGCGCTA 2198
DB 2136 CAAGATGCTGCTTA 2150

RESULT 9
AX082544
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1. 2151
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="mutant from Rhodotorula glutinis"

Query Match 49.54; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.04; Pred. No. 2.5e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 GCGCGCTCCAGCAAGTGGCTGCGGCGCCGCACTTGGCTCTGCGCGCGCGCGCT 158
DB 45 CGTGCATCCGCAAGAGAGGCTGTCATGCGCGCTGACCAACTGCGAGTGCAGGCTC 104
QY 159 CGATGGCCAGCG 218
DB 105 GCACCTCCCAACACCGAGTCAAGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 164
QY 219 CACCGACAGCTGCTGAGTCAAGCGGGAACAGCTCAACCTCAACCTCAACCTCAACCTCAAC 278
DB 165 GACCGATCCAGCGGTGAACTGACGGCTAATGCTCAACCTGCGAGAGTGTGCTCGCG 224
QY 279 GCGCCGCAAGGCG 338
DB 225 GCGAGAGAGGAGGAGGCTGTGCGGTCAAGGACAGCGAGATCCGCTCAAGATTGA 284
QY 339 CAAGAGCTGACCTTCTCAAGGCGCGCTTCAAGCTGCTGAGATCGCTCAAGAGTCAACACG 398
DB 285 CAATGCTGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 344
QY 399 TTTGCTGCTGCG 458
DB 345 ATTGGGGAATCCGACACACCGCACCGAGGACGCGCATGCTGCTCAAGAGCTCTCTCT 404
QY 459 CGACGACAGCTGCTGCG 518
DB 405 CGACGACAGCTGCTGCG 464
QY 519 CTTGAGAACAGCTTCTGCTGAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 578
DB 465 TCTGAGAACCTGCTTCTGCTGAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 524
QY 579 GCTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
DB 525 CTGACCTCGGCG 584
QY 639 GAACCAACGATCAAGCGCTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 698
DB 585 CAACCAACGATCAAGCGCTTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 644
QY 699 CAGCCGCTCTGTAATGCG 758

ORIGIN

Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.5e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY CGCGCGCTCGACCAACTCGGCTCGGCGCCCACTTCGAGCTTCGCGCGACCGCCGCT 158
DB CGTCGATCCGCAAAACAGAGCTGTCAATGCGGCTTCGACCAACCTCGAGTCGAGCTC 104
QY 159 CGATGACCAAGCGCGGCAACCACTGCAAGTGTGCAAGAGCTCTCAAGCAACC 218
DB 105 GCACTGCGCCCAACCAACAGAGTCAAGAGTCACTGCAAGAAAGTCTCGCGCGCC 164
QY 219 CACCGACAGAGTGTGAGTCAAGAGGAGTACAGCTCAAGCTCGAGTGTGAGCTG 278
DB 165 GACCGACTCGAGGCTGAACTGACGGCTACCTCGCTCAACTCGAAGACGTCGTGCGC 224
QY 279 CGCGCCGAAGGCGCGAGGCTCGCTCGCAAGACGACGAGATCGCGGCAAGCTCGA 338
DB 225 CGGAGGAAAGGCGAGGCTGTCCGCTCAAGGACAGCGACGAGATCGGCTCAAGATTGA 284
QY 339 CAAGAGCTGCACTTCCTCAAGGCGGAGCTTCAGAACTCGGCTCAAGAGTCAACAGG 398
DB 285 CAATGCTGAGTCTTGTTCGCTCGCAACTCTCCATGAGGCTCAAGGCTCAAGACTGG 344
QY 399 TTTCGGTGGCTCGGCGGACAGAGAGTGAAGATGCACTCAAGCTCGAAGGCGCTCAT 458
DB 345 ATTGGCGGATCCGAGAGACCCGACAGAGAGGAGCCATCTCGCTCCAGAAAGGCTCTCT 404
QY 459 CGAGCACCAAGCTCTGCGGCTGACCCGACGTCGCTCTCTTCAAGCTTCGAGCGG 518
DB 405 CGAGCACCAAGCTCTGCGGCTGACCCGACGTCGCTCTCTTCAAGCTTCGAGCGG 464
QY 519 CTTGAGAACAGCTTCGCTCGAGTGTGTCGCGCGGCGCATGTCATTCGCGCTCAATC 578
DB 465 TCTCGAAGACTGCTTCCTTCGAGTGTTCGCGCGGCGCATGTCATTCGCGCTCAATC 524
QY 579 GCTCAAGCTGCGCACTCGGCGGCTCGGCTGCTGCTCTTGAAGGCGCTCAACATCTT 638
DB 525 CTTGACCGCGGCGCACTCGGCGGCTCGGCTGCTGCTCTGAGGCGCTCAACATCTT 584
QY 639 GAACCAAGGATCAAGGCGCATGTCGCTCGGCGGCTTCATTCGCGGCTCGGCGCAT 698
DB 585 CAACCAAGGATCAAGGCGCATGTCGCTCGGCGGCTTCATTCGCGGCTCGGCGCAT 644
QY 699 CAGCGCGCTCTGTCATTCGCGGCGGCGCATTCAGCTCAACCGCACTCAAGGTTTCACT 758
DB 645 CTCTCTCTCTCTCAATTGAGCGGCGCATTCAGGCGTCAACCGCACTCAAGGTTTCACT 704
QY 759 TTTGACAGAGGAAACCGAAGATATATTTTTCGCGGAGGCACTTCGCTCTTTGCTT 818
DB 705 CGTTCACAGAGGAAACCGAAGATATCTGTACGCGCGGAGCGCATTCGCTCTTCAACT 764
QY 819 CGAGCGAGTGTCTCGGCGCGGAGAGAGTGTCTGCTGTGTCAGGAAAGGCGGCTC 878
DB 765 CGAGCGGCTGTCTCGGCGCGGAGAGAGTGTCTGCTGTGTCAGGAAAGGCGGCTC 824
QY 879 CGCTCGATGCGACCTTCAGTCTTCAGACTTCGACATGCTCTCGCTCTCTCGAGG 938
DB 825 ACGATTCGATGCGACCTTCAGTCTTCAGACTTCGACATGCTCTCGCTCTCTCGAGT 884
QY 939 CTTGAGCGGCTTCAGAGTGAAGGAGCATGTCGCGGAGGCGTCTTCGCGGCTTCAT 998
DB 885 GCTTCAGGCGATGAGGTCGAAAGGATGTGCGCAAGCGGCTCTTCACCTCTTCT 944
QY 999 CCACGAGCTTCGCGCGCGCAACCGCGGCAAGTGAAGTGCAGGCGCAATCCGCAAGCT 1058
DB 945 TCAGACGTCAGCGGCGCTCAACCGAGCAAGTGAAGTGCAGGAAAGTTCGCAAGCT 1004
QY 1059 CTTTCGAGCTGTGTTGCGCTTGAAGACGAGAGAGAGTCAAGTCAAGACGACGA 1118
DB 1005 CTTGAGGAGAAAGCGCTTGTCTTCACCATGAGAGAGAGTCAAGTCAAGACGACGA 1064

QY 1119 GGGCATCTTGCCAGGAGCGCTACCGGCTCCGAGCTGCGCTCAAGTTCCTCGGCGCGCT 1178
DB 1065 GGGCATCTTGCCAGGAGCGCTACCGGCTCCGAGCTGCGCTCAAGTTCCTCGGCGCGCT 1124
QY 1179 CGTGAGGACATGATGACGCTTACTGACTCTCTCGCTCGAAGAACAA--CGACGAC 1235
DB 1125 CGTCAGGACCTCATTCACGCGCCACGCGCTCTTACCATGAGAGCGGCGGACATGACGAC 1184
QY 1235 CGACACCGGCTCTCTGACGTCGAGAACAGCAGACCGGCGGACGCTTCCAGG 1295
DB 1185 CGACACCGGCTCTCTGACGTCGAGAACAGCAGCTTCGACACGCGGCGGCAATTCAGG 1244
QY 1296 GTGCGCTGTCTGATTCGATTCGAGAACACAGGCTTCGCTCGCTTCATCGGCAAGCT 1355
DB 1245 TGCGGCTGTGCGCAACACATGAGAGAACCTCGCTCGGCTCGGCGGCAAGT 1304
QY 1356 CACTTCAGGAGTGCACCGAGTGTCAACGCTGCTCAATGACCGGCGCTGCTTCG 1415
DB 1305 CACTTCAGGAGTGCACCGAGTGTCAACGCGGATGAAACCGGCGCTTCTCTCTG 1364
QY 1416 CCGGCTGCGAGGAGCGGCTCGCTCAACTATCAAGGAGGCTTGAACATTCACATCG 1475
DB 1365 CTTGCGGCGGAAACCGCTCGCTCTCTCTCACTGACAGAGGCTTGAATGCGCTG 1424
QY 1476 TGCTTACGCTTCGAGCTCGGCAACTTGCACACCGGCTCACTTCTGTCAGCGCG 1535
DB 1425 GAGCTACACCTCGAGTTGGAACACTCGCAACCTCTGACACGACATGTCCAGCGCG 1484
QY 1536 AGAGTGGGACCAAGGCGCTCACTGCTGCTGCTCACTTCGCGCGCGGCACTGCGCA 1595
DB 1485 TGAAGTGCAGAACGAGGCTGCACTGCTGCTCACTTCGCTGCTGCGACGACGCA 1544
QY 1596 GGGCAACGAGCTCTTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
DB 1545 GTCCAGACGCTCTTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
QY 1656 GAGCTTCGCGCGAGTGAAGTCTCACTTCAAGAACAGTTCGAGCGGCTTCTGCGGCT 1715
DB 1605 GAGCTTCGCGCGAGTGAAGTCTCACTTCAAGAACAGTTCGAGCGGCTTCTGCGGCT 1664
QY 1716 CTTGAGAGGAGCTGCGGCACTGAGCTGAGGCA-----AGCACTTTCGCT 1763
DB 1665 CATGACGAGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1724
QY 1764 CAGAGTGAAGAGGCGTCAACAGGCTTCGAGAGAGAGAGAGAGTCAAGCTTCGAGG 1823
DB 1725 GAGGTGAACAGAGGCTTCGCAAGGCGCTCGAGAGAGAGAGAGTCAAGCTTCGAGG 1784
QY 1824 GCGCTGCGAGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1883
DB 1785 GCGCTGCGAGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1844
QY 1884 GCGCTTGCAGAGCTCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1943
DB 1845 GTC-----GCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1895
QY 1944 GGGCATTCGCTCAAGGAGTGCAGACCGCTTTCGAGAGAGCGCTTTCGAGG 2003
DB 1896 GGGCATTCGCTCAAGGAGTGCAGACCGCTTTCGAGAGAGCGCTTTCGAGG 1955
QY 2004 GCGGCGGAGAGTCTCTGCGCGGAGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2063
DB 1956 GCGGCGGAGAGTCTCTGCGCGGAGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015
QY 2064 GCTTCGAGGAGAGGCGGCGGAGAGGAGTGTTCGAGGAGAGAGAGAGAGAGAGTTCG 2123
DB 2016 GCTTCGAGGAGAGGCGGCGGAGAGGAGTGTTCGAGGAGAGAGAGAGAGAGTTCG 2075
QY 2124 GAGGAGAGTCTGCGGATTCAGAGGAGTTCAGAGAGGCGGAGTTCAGAGAGTTCG 2183
DB 2076 CTGCAAGCTCTCAAGATTCAGAGGAGTTCAGAGGAGTTCAGAGGAGTTCAGAGGAGTTC 2135

2y 2184 CAAGATGCTGCGCTA 2198
2b 2136 CAAGATGCTGCTTA 2150

RESULT 11
LOCUS RPPALGNA 2315 bp mRNA linear PLN 23-MAR-1993
DEFINITION R.tornuloides mRNA for L-phenylalanine ammonia-lyase (PAL).
ACCESSION X12702.1 GI:288350
VERSION X12702.1
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
REFERENCE 1 (bases 1 to 2315)
AUTHORS Unknown.
JOURNAL Unpublished
COMMENT (Patented) BP 0 260 919 A1.
FEATURES
location/Qualifiers
1..2315
/organism="Rhodospiridium toruloides"
/mol_type="mRNA"
/strain="IFO 559"
/db_xref="taxon:5286"
/clone="psw13"
1..2315
/gene="PAL"
1..2151
/gene="PAL"
/codon_start=1
/product="L-phenylalanine ammonia-lyase"
/protein_id="CA31209.1"
/db_xref="GI:288351"
/db_xref="GOA:P11544"
/db_xref="SWISS-PROT:P11544"
/translation="MAPSLDISISHSFANGVAKAKQAVNGASINLAVGSHLPTQVYQ
VDIVKMLAAPTDSLEIDGSLMGDVVAKRGKRVKQSDERKIDKSEVELR
PLEVNGVYGVTFGGSGADTEDEAISIQKALHEIQCGVLPSPSPFGRGLENLPI
SMATLALHDAMLSLSOSLTFMTVEAMVGHAGSPHFPLDTPRPHPTQEVAGNRK
LLEGRFVHHBEVKVXDDEGILKQRPYRTSPQMLGTVSDLIHAHAYVTTEBQ
STTNPLLDVENKTSHGNGNQAAAVANTKEKTLGLAQIKLFTQUTBMLAGNR
GLPSCLAEDPSLSYHCKGLDIAAAVATSEIHLAVITHVQVAPAMANOAVNSLMI
SARITBENDVLSLLATHLKCYQALIDRLAIEFEKPKGPALVSLIDQHFQSAVTG
SNRHEIVKXNKLAKRIEQTNGYDAIDPRMHDFAFGATGVEVLSSTLSIAVNA
WKVAABSAISLTROVRETFMSASTSPALSYLSPRTQILYAFVBEELGVAKRBDV
PLGKEVITGSNISKIYBAIKSGRINNVLLKMLA"
2315
/gene="PAL"

polYA_site
/gene="PAL"

Query Match 49.5%; Score 1196.6; DB 8; Length 2315;
Best Local Similarity 74.0%; Pred. No. 2.4e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

2y 99 CCGCGCTCCGACCAAGTCGCGTGCAGGCGCCCACTTCGCTCCGCGCACGCGCGCT 158
2b 45 CCGTCGATCCGAAAGCAAGGCTGTCAATGCGCCTCGACCACTCGACGCGAGGCTC 104
Qy 159 CGATGGCCACCGCGCGCACCACTCGCACTGCAATGCTGCGAGAGAGCTCTCAAGCACCC 218
Db 105 GCACTGCGCCCAACCCAGGTCACGCAAGTCACTGTGGAAGAAGTGTCTGCGCGGCGC 164
Qy 219 CACCGACGACGTCGTGCACTCAGCGGCTACAGCTCACTCGCTGCACTGTCGCGCG 278
Db 165 GACCTGACTGAGCGCTGCACTGAGCGCTACTCGCTCACTCGAGAGCGTCTGCTCGCG 224
Qy 279 CGCCCGCAAGGGGCGCAGGGTCCGCTGCAAGACGACGACGAGATCCGCGCACCGCTCGA 338
Db 225 CGGAGAGAGGCGCGCTGTCCGCTCAAGGACGACGACGAGATCCGCTCAAGATTTGA 284

Qy 339 CAAGAGCGTCGACTTCCTCAAGGCCAGCTTGAAGACTCGATCTACGAGATGACACAGCG 398
Db 285 CAATCGGTGAGATTCCTGCGCTCGCACTCTCCATGAGCGCTACGGCTGACAGCTGG 344
Qy 399 TTTCGTGCTCGCGCGACACGAGACTGAGATGCACTGACCTCTCAAGAGCGCTCAT 458
Db 345 ATTGGCGGATTCGCGAGACACCGCACCGAGACCGCATCTGCTCCAAAGGCTCTCT 404
Qy 459 CGAGCACCACTCGCGCGGCTGACGCGGACGTCGCTGCTGCTTACGCGTGGACGCG 518
Db 405 CGAGCACCACTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 464
Qy 519 CTTGAGAACACGCTTCGCTGAGGTGCTGCGGCGCGCATGATGATTCGCTGCACTC 578
Db 465 TCTGAGAACCTCGCTTCCTTCGAGGTGCTGCGGCGCGCATGATGATTCGCTGCACTC 524
Qy 579 GCTCAGCGGTGCGACCTCGCGCGCTCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCT 638
Db 525 CTTGACCGCGGCGACCTCGGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
Qy 639 GAACCAACCGATCAGCGCCATCGTCCCTCCGCGGCTCATCTCGGCGTGGGCGACCT 698
Db 585 CAACCAACCGATCAGCGCCATCGTCCCTCCGCGGCGCATCTGCGTGGGCGACCT 644
Qy 699 CAGCGCGCTCTGTAATGCGCGCGCGCATCAGCGCTCAACCGGAGCTGAAGTTACGT 758
Db 645 CTTCTCTCTCTCTCACTTGACAGCGGCGCATCAGCGCTCAACCGGAGCTGAAGTTACGT 704
Qy 759 TTGACAGAGGAAACGAGAAGATCATGTTTGCGCGGAGGCGCATCTGCTCTTGGTCT 818
Db 705 CTTCAAGAGGAAACGAGAAGATCTGTAAGCGCGCGGAGCGATGCGCTCTTCAACCT 764
Qy 819 CGAGCGATGCTCTGCGCGCGCGGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 878
Db 765 CGAGCGCGTCTCTGCGCGCGCGGAGAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 824
Qy 879 CGCTCGATGCG 928
Db 825 AGCATCATGTCACCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 884
Qy 939 CTTACGCGCTCTCAAGGTGAGGCGCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCT 998
Db 885 GCTCAGCGCATGACGCTGCAAGGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 944
Qy 999 CGAGCGCTCTGCT 1058
Db 945 TCAGGACCTCAAGCG 1004
Qy 1059 CTTTCGCGCTCGTCTGCGCTTGAAGCAAGAGAGGTCAAGGTCAAGAGCGACGA 1118
Db 1005 CTTGAGGAGAGCGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
Qy 1119 GGGCATTTTTCGCGAGACCGGTAACCGGCTCGCGCGCTGCGCTGCTGCTGCGCGCGCT 1178
Db 1065 GGGCATTTTTCGCGAGACCGGTAACCGGCTTTCGCGCGCTGCTGCTGCTGCTGCGCGCGCT 1124
Qy 1179 CGTGAAGACATGATGACGCGCTCTGCACTCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db 1125 CTTGAGGACCTCAATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1184
Qy 1236 CGACACCGCGCTCTGCGAGCTGAGAACAGACGCGCGCGCGCGCGCGCGCGCGCGCG 1295
Db 1185 CGACACCGCGCTCTGAGCGCTGAGAACAGATTGCGACACGCGCGCGCGCGCGCGCG 1244
Qy 1296 GTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
Db 1245 TCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
Qy 1356 CAACCTTCAAGAGTCAAGAGTGTCTCAAGCTGCTCAAGCTGCTGCTGCTGCTGCTGCT 1415
Db 1305 CAACCTTCAAGAGTCAAGAGTGTCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364

Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CGCGCTCCGACCAAGTCGGCTGCGGGCCCACTTCGGCTCTCGCGCAGCCCGGCT 158
 Db 45 CGTCGATCCGAAAGCAGGTGTCAATGCGGCTCGACCAACCTCGAAGTCGAGCTC 104
 QY 159 CGATGGCCACGCGCGCCACCAAGTCGAGCTGAGATTCGTGAGAGACTCTCAAGCACC 218
 Db 105 GCACTGCCCAACCAACCAAGTCGAGCTGAGATTCGTGAGAGAGACTCTCGCGGCC 164
 QY 219 CACCGACGACGTCTGAGCTGAGGGGTAACAGCTCAGCTCCGCTGAGCTGCGCGC 278
 Db 165 GACCGACTGAGAGCTGAACTGAGCGCTACTCGCTCAACTCGGAAGCTCTGCTGGC 224
 QY 279 CGCGCGCAAGGGGCGCAAGGTCGCGCTCAAGACGACGAGATCCGCGCAGCGTGA 338
 Db 225 CGCGAGGAAGGGCGCAAGGCTCGCGCTCAAGACGAGATCCGCTCAAGATTGA 284
 QY 339 CAAGAGCGTGACTTCCTCAAGGCCCACTTCAGAACTCGCTTACAGGATCAACAGGG 398
 Db 285 CAATGCTGAGTTCCTTCGCTCGCACTCTCAAGAGCTTACGCGCTCAAGCTGG 344
 QY 399 TTTGCTGCTCGCGCGCACAGAGACTGAGATGAGTCACTCAAGCTCCAGAGGCGCTAT 458
 Db 345 ATTTGGCGGATCCGACAGACCCCGACAGAGAGCCATTCGCTCCAGAAAGCTCTCT 404
 QY 459 CGAGCACAGCTCTGCGCGCTGACCCGACGTCGCTCTTCAAGCTCGAGCGCG 518
 Db 405 CGAGCACAGCTCTGCGCGCTGACCCCTTCTGCTGACCTGCTCCGCTCGCGCGCG 464
 QY 519 CCTCGAAGACAGCTTCGCTCGAGGTGTCGCGCGCGCATGGTCAATCGCGCTCAACTC 578
 Db 465 TCTCGAAGACTGCTTCCCTCGAGTGTTCGCGCGCGCATGCAATCGCGCTCAACAG 524
 QY 579 GCTCAAGCGTGGCACTCGCGCTGCTCGCTGCTTGAAGCGCTCAACAATCTT 638
 Db 525 CTTCGACCGCGCGCACTCGCGCTGCTCGCTGCTCGAGCGCTCAACAATCTT 584
 QY 639 GAACACCGCATCAAGCCCATGTCGCCCTCGCGCGCATGCTCGCGCGTCCGCGCATCT 698
 Db 585 CAACCAAGCATCAACCCCATGTCGCCCTCGCGCGCATCAATCTTCGTCGCGCATCT 644
 QY 699 CAGCCGCTCTGTCATGCGCGCGCATCAACCGCATCAACCGCATCAAGTTCATCT 758
 Db 645 CTCTCTCTCTCTCAATTCGACAGGGCGCATCAAGGCTCAACCGCATCAAGTTCATCT 704
 QY 759 TTTGACAGAGGAACCGAAGATGATGTTTGGCGCGAGAGCCATCTCTTTTGTCT 818
 Db 705 CGTCCACAGAGGCAAGAGAGATCTGTAGCGCGCGAGCGCATGCGCTCTTCAACT 764
 QY 819 CGAGCAGTCTCTCGAGCGCGAAGAGGCTCTGCTGATCAACGGAAGCGCGCTCTC 878
 Db 765 CGAGCCGCTCTCTCGAGCGCGAAGAGGCTCTGCTGATCAACGGAAGCGCGCTCTC 824
 QY 879 CGCTCTGATGCGAGCTCTCACTGTCAGACTGCGACATCTCTCTCTCTCTCGAGGC 938
 Db 825 AGCATGCTATGCGAGCTCTCTCTGACGAGACGACATCTCTCTCTCTCTCTCGAGTC 884
 QY 939 CTTCAGCGCTCTCAAGGTCAGAGGTCAGTTCGCGACAGAGGCTCTCTCTCTCTCTCT 998
 Db 885 GCTTCAGCGCTCTCAAGGTCAGAGGTCAGTTCGCGACAGAGGCTCTCTCTCTCTCTCT 944
 QY 999 CCACAGCTCTGCGCGCGCGACCGCGCGAGTTCAGAGTTCGCGCGCAATCCGCGAGCT 1058
 Db 945 TCACAGCTCTGCGCGCGCGTCAACCGAGAGTTCAGAGTTCGCGCGAATCCGCGAGCT 1004
 QY 1059 CTTTCGCGCT 1118
 Db 1005 CTTTCAGAGGAAGCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1064
 QY 1119 GGGCATTTCTGCGAGGACGCTACCGCTCCGACGTCGCTCAGTCTCTCTCTCTCTCTCT 1178
 Db 1065 GGGCATTTCTGCGAGGACGCTACCGCTCCGACGCTCTCTCTCTCTCTCTCTCTCTCT 1124

QY 1179 CGTGGAGACATGATGACAGCTTACTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1235
 Db 1125 CGTACGACACTCTATTTACCGCCCAACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1184
 QY 1236 CGACACCCGCT 1295
 Db 1185 CGACACCCGCT 1244
 QY 1296 GTGCGCT 1355
 Db 1245 TCGCGCT 1304
 QY 1356 CACTTCAAGGTCAGCGAGTTCCTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1415
 Db 1305 CACTTCAAGGTCAGCGAGTTCCTCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1364
 QY 1416 CCTGCTGCGAGGACCGCT 1475
 Db 1385 CTTGCGCGCGAAGACCT 1424
 QY 1476 TGTTCAGCTTCGAGCTCGAGCACTTTCGCAACCGGATCACTTCTCTCTCTCTCTCTCT 1535
 Db 1425 GGGCTACACCTCGAGTTGGGACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1484
 QY 1536 AGAATGGGCAACGAGGCGTCAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1595
 Db 1485 TGAATGGGCAACGAGGCGTCAACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1544
 QY 1596 GGCCACGACGTCCTTCT 1655
 Db 1545 GTCCACGAGCTCTTCT 1604
 QY 1656 CGACTCTCGCGCATGAGCTGACTTCAAGAGAGTTCGACCGCTCTCTCTCTCTCTCTCTCT 1715
 Db 1605 CGACTCTCGCGCATGAGCTGACTTCAAGAGAGTTCGACCGCTCTCTCTCTCTCTCTCTCT 1664
 QY 1716 CTTCCAGACGACCTGCGGCACTGCGGCTCGACCGCA-----AGCAGCTTGGCGT 1763
 Db 1665 CATGACCGACCTTGTGCT 1724
 QY 1764 CGAGTCAAGAGGCGCTCAACAGGCTCTCGAGCAGACGAGCAGTACGATCGATCGAGCC 1823
 Db 1725 GAGGTGAAGAGCGCTCTCGCAAGCGCTCGAGGAGACCAACTGATGAGCTCTGCTCC 1784
 QY 1824 GCGCTGCGACGAGCGCT 1883
 Db 1785 GCGCTGCGACGAGCGCT 1844
 QY 1884 GCGCTTGCACAGTCAACCTTACTGCGGTCACAGCGGTGAAGTTCCTCGCGAGAA 1943
 Db 1845 GTC-----GCTCTGCTCTGCGCGCGCTCTCAACGCTCTGAAGGTCGCGCGCGAGTTC 1895
 QY 1944 GGCATTTGCTCTACGCGCGAGGTGCGCAACCGCTTCTGCGAGACGCGCTTCTGCGAGCC 2003
 Db 1896 GGCATTTGCTCTACGCGCGAGGTGCGCAACCGCTTCTGCGAGACGCGCTTCTGCGAGCC 1955
 QY 2004 GCGGCGGACGACATCT 2063
 Db 1956 GCGGCGCTCTGTACT 2015
 QY 2064 GCTCGGCTGACAGCGCGCGCGCGAGTGTCTGCTGCGCTGCGAGCAGAGACGATCGG 2123
 Db 2015 GCTTGGGCTCAAGCGCGCGCGCGAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2075
 QY 2124 GAGCAAGTCTGCGGATCTAGAGGCTCTCAAGAGCGCGCGCATCAACAGTCTCTCTCT 2183
 Db 2076 CTCGAGTCTTCCAGATCTTACAGGCTCAAGGCTCAAGGCTGCGGAGATCAACAGTCTCTCT 2135
 QY 2184 CAAGATGCTGCGTA 2198
 Db 2136 CAAGATGCTGCTTA 2150

RESULT 14
LOCUS E01785 2331 bp DNA linear PAT 28-SEP-1997
DEFINITION DNA sequence coding for L-phenylalanine ammonia-lyase.
ACCESSION E01785
VERSION E01785.1 GI:2170038
KEYWORDS JP 1988317088-A/1.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
1 (bases 1 to 2331)
Fukuhara, N., Yoshino, S., Watanabe, M., Yamamoto, K., Suzuki, M. and
Nakajima, Y.
CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
Patent: JP 198317088-A 1 26-DEC-1988;
JOURNAL MITSUI TOATSU CHEM INC
COMMENT OS Rhodospiridium toruloides
PN CP 1988317088-A/1
PD 26-DEC-1988
PF 18-JUN-1987 JP 1987152359
PI FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
YAMAMOTO KAORU,
PI SUZUKI MAKI, NAKAJIMA YOSHIYUKI
PC C12N15/00,C12N9/88,(C12N9/88,C12R1:01);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=IFO 559 (ATCC 10788);
FH key Location/Qualifiers
FT
FT CDS 1..2151
FT /product='L-phenylalanine ammonia-lyase' FT
3'UTR 2152..2331.
Location/Qualifiers
1..2331
/organism='Rhodospiridium toruloides'
/mol_type='genomic DNA'
/db_xref='taxon:5286'
FEATURES
source
ORIGIN
Query Match 49.5%; Score 1196.6; DB 6; Length 2331;
Best Local Similarity 74.0%; Pred. No. 2.4e-136;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
QY 99 GCGCGCTCCGACCAAGTGGCGTGGGCGGCCGCACTTCGCGCTCTCGCGCGACGCGCGCGCT 158
Db 45 CTTCCATCCGTAAGCGAGGCTGTCAATGGCGCTTCAACCACTTCGAGTCCGAGGCTC 104
QY 159 CGATGGCCACGCGCGCGACCAAGTCGACGCTTGAGATCGTGCAGAGCTCTTCAAGCC 218
Db 105 GCACTGCCCAACCAACCGAGTCACGACGTCGACATGTCGAGAAAGATGCTCGCGCGGC 164
QY 219 CACCGACGACGTCGTGAGCTTACGAGGCTACAGCTTCCGTACGATGTTGTGCGCGC 278
Db 165 GACCACTCGACGCTCGAAGTTCGAGGCTTACTGCTCAACTCGAGACGTCGTGTCG 224
QY 279 GCGCGCGAAGGGGCGCAGGCTCGCGCTCCAGAAAGACGACGAGATCGCGCGACGCTCGA 338
Db 225 CCGGAGAAAGGCGAGGCTGTCCGCTTAAGAGACGACGAGATCCGCTCAAGATGA 284
QY 339 CAAGAGCGTGAATTCTTCAAGGCCCACTTCAAGACTCGTCTTACGAGAGTCACACG 398
Db 285 CAATTCGCTGAGATCTTCCGCTCGCACTTCCATGAGCGTCTACGCGCTCAAGATGG 344
QY 399 TTTCGCTGCTGCGCGCGACGAGAGCTGAGAGATGAGTGCAGCTCCAGAGAGCGCTCAT 458
Db 345 ATTTGGCGATCGAGACACCGGACGAGAGCGACATTCGCTCCAGAGAGCTTCTCT 404
QY 459 CGAGCACCAAGCTCTGCGCGTGACCGCAGACGTCGATCGATCTTCAAGCTCGGACGCG 518

Db 405 CGAGCACCAAGCTCTGCGCGTGTCTCCCTTCGTGCTGACTCGTCCGCTCGGCGCGG 464
QY 519 CTTGAGAAACAGCTTCCGCTTCGAGGTCGTCGCGGCGGCGCATGATCGGTCAGCTC 578
Db 465 TTTGAGAACTCGGTTCCCTTCGAGGTTGTCGCGGCGGCGCATGACATCGGCTCAACG 524
QY 579 GCTCAGCGCTG3CCACTCGGCGGTCGCGCTCTGCTCTTGAAGCGCTCAACCACTTCT 638
Db 525 CTTGACCGGCGGCGCACTCGGCTGTCCGCGCTGTGCTCTGAGGCGCTCAACCACTTCT 584
QY 639 GAACCAACCGATACCGGCGCATGCTCCCTCGGCGGCTCACTTCGCGCTCGGCGCACT 698
Db 585 CAACCAACCGATACCGGCGCATGCTCCCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 644
QY 699 CAGCGCGCTCTGACATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 758
Db 645 CTTCTCTCTCTCTCAATTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 704
QY 759 TTGACAGAGGAAACGAGAAATCATGTTTGCAGCGAGGCGCATCTCGCTTTTGTGCT 818
Db 705 CGTCCAGAGGCGCAAGAGAAATCTGTAAGCGCGCGGCGGCGGCGGCGGCGGCGGCG 764
QY 819 CGAGGCACTGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 878
Db 765 CGAGCGGCTGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 824
QY 879 CGCCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 938
Db 825 AGCATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 884
QY 939 CTTGACGCTCTCAAGTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 998
Db 885 GCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 944
QY 999 CCAGCAGCTGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1058
Db 945 TCAAGACGTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1004
QY 1059 CTTTCGCGCTGCTGTTTGGCGGTTGAGACGAGAGAGAGGTCAGGTCAGAGACGAG 1118
Db 1005 CTTGAGAGGAAAGCGGCTTGTGTCACATGAGAGAGGTCAGGTCAGAGACGAG 1064
QY 1119 GGGCATTTTCCGAGAGACGCTACCGGCTCGGAGGTCGCTCAGTTCGCGCGCGCT 1178
Db 1065 GGGCATTTTCCGAGAGACGCTACCGGCTCGGAGGTCGCTCAGTTCGCGCGCGCT 1124
QY 1179 CGTGGAGACATGATGACGCGCTACTGACTCTCTGCTGAGAACAA--CGAGAC 1235
Db 1125 CGTACGACCTCATTTACGCGCGCGGCGGCTCTACCATGAGGCGGCGGCGGCGGAC 1184
QY 1236 CGACACCGGCTCTCTCGACGTCGAGAACAGACAGACCGGCGGCGGCGGCGGCGG 1295
Db 1185 CGACACCGGCTCTCTCGACGTCGAGAACAGACCTTCCACACGCGGCGGCGGCGGCG 1244
QY 1296 GTGCGCTGCTCGATTTGATGAGAGACGAGGTCGACGCTCGCTTCAAGAGCT 1355
Db 1245 TGCGCTGTGCGCAACACATGAGAGACCTCGCTCGGCTCGGCGGCGGCGGCGG 1304
QY 1356 CAATTTCAGAGTGCACCGAGTTGCTCAAGCTGCAATGAACCGGCGGCTGCTTGTG 1415
Db 1305 CAATTTCAGAGTGCACCGAGTGTCAAGCGCGGCGGCGGCGGCGGCGGCGGCGG 1364
QY 1416 CCGCGCGGCGGAGACCGGCTGCTCAACTATACGAGAGGCGGCTGAGATTCAGATCG 1475
Db 1365 CCGCGCGGCGGAGACCGGCTGCTTCTTACACTGAGAGGCGGCTGAGATTCAGATCG 1424
QY 1476 TGCTTACGCTTCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1535
Db 1425 GAGCGTACGCTCGAGGTTGGAGACCTTCGCGGCGGCGGCGGCGGCGGCGGCGG 1484
QY 1536 AGAGTGGGCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1595

Db 1125 GGTCAAGGACCTCATTTCAAGCCCAAGCCCTCTCACCATCGAGGCCGCGCATGTGACGAC 1184

2Y 1236 CGACAAACCCGCTCTCGACGTGAGAAACAAGACAGCGCGACGGCGCACTTCCAGGC 1295

Db 1185 CGACAAACCCCTCTCATCGACGTGAGAAACAAGACTTCCACACACGGCGCAATTTCCAGGC 1244

2Y 1296 GTGGCTGTCTCGATTTCGATGAGAAAGACAGGCTCGCACTGGCCCTCATTCGGCAAGCT 1355

Db 1245 TCCGCTGTGGCAACAACCATGAGAAAGACTCGCTTGGGCTCGCCCAAGTCGGCAAGCT 1304

2Y 1356 CAACCTTCAGCAGTAGCAACCGAGTTGCTCAACGTGACATGAAACCGGCGCTGCTTCTGTG 1415

Db 1305 CAACCTTCAGCAGTAGCAACCGAGTTGCTCAACGTGACATGAAACCGGCGCTTCCCTCTGTG 1364

2Y 1416 CCTGCTGCCAGGACCCGCTGCTCAACTATACAGGCAAGGCTTGGACATTCACATCCG 1475

Db 1365 CCTCGCGGCCAGAACCCCTCGCTCTCTCACTGCAAGGCGCTCGACATCGCCGCTCG 1424

2Y 1476 TGCTTAAGCTTCGGAGGCTGGGCACTTGGCAACCGGCTCACTACCTTGTCCAGCCCGC 1535

Db 1425 GAGCTACACCTCGAGATTGGGACACCTCGCAACCTGTGACAGCATGTCCAGCCGCG 1484

2Y 1536 AGAGATGGGCAACCAAGCCGTCAACTGCTGCTCATCTCCGCGCGCGCACTGCCGA 1595

Db 1485 TGAGATGGCGAACAAGGCGGTCAACTGCTTGCCTCATCTCGGCTGTGCAACGCA 1544

2Y 1596 GGCACAAGACGTCTTTCTTCT 1655

Db 1545 GTCCACACAGCTCTCTTCT 1604

2Y 1656 CGACCTCCGCGGATGAGAGCTGCACTTCAAGAACAGTTCGACCCGCTTCTCCGACTCT 1715

Db 1605 CGACTTGGCGCGATCGAGTTGAGTTCAAGAACAGTTCGACCCGCTTCTCTCTCTCTCT 1664

2Y 1716 CCTCCAGAGCACTCGGCACTGGCTCTGACGTCA-----ACGCACTTGGGCT 1763

Db 1665 CATGACCAGACATTTGGCTCCGCATGACCGGCTCGACCTGCGCAACGAGCTCTGTCGA 1724

2Y 1764 CGAGGTCAAGAGAGGCGTCAACAAGCGTCTGAGCAGACGACGATGACGATCTCGAGCC 1823

Db 1725 GAAAGTGAACAAGAGCTTCGCGCAAGCGCTTGAAGACACAACTGTACGACCTCTCTCC 1784

2Y 1824 GCGCTGAGCAGACGCTTCTGTACGACACCGGACCGCTGCTGAGCTCTCTCTCTCTCT 1883

Db 1785 GCGCTGGACAGACGCTTCT 1844

2Y 1884 GCGCTCTGCAACGTACCTTACTGCGCTTCAACGCGTGAAGGTTGCTCTGGCCGAGAA 1943

Db 1845 GTC-----GCTCTGTCTGCGCGCGTCAACGCTGGAAGTTCGCGCGCGAGTC 1895

2Y 1944 GGCATCTCGCTCAAGCGAGAGGTGCGCAACGCTTCTGCGAGACGCGGCTTCTGCGAGGC 2003

Db 1896 GGCATCTCTGCTCAACCGCGCAAGTCCGAGACCTTCTGTCTCGCGCGGTGACCTCTCT 1955

2Y 2004 GCGGCGCAGCAGCATCTCTGCGCGCGCAGCGCGCTCTCTCTCTCTCTCTCTCTCTCT 2063

Db 1956 GCGCGCGCTCTGTACTCTCTGCGCGCGCATCAGATCCTTCAACGCTTCTCTCTCTCTCT 2015

2Y 2064 GCTCGCGCTGAGAGCGCGCGCGCGCGAGCTTGTCTGCGCTGAGCAGAGAGAGATCGG 2123

Db 2016 GCTTGGCGTCAAGGCGCGCGCGCGAGAGCTTCTCTGCGCAAGAGGTAAGATCGG 2075

2Y 2124 GAGCAACGCTGCGCATCTCAAGGCGCATCAAGACCGCGCATCAACGATCTCTCTCT 2183

Db 2076 CTCGAACGCTCTCAAGATCTTCAAGGCGCATCAAGTCCGCGAGATCAACACGCTCTCT 2135

2Y 2184 CAAGATGCTCGCGTA 2198

Db 2136 CAAGATGCTCGCTTA 2150

CC host cell comprising the polynucleotide is useful for the production of L
CC -PAL, by adding the polynucleotide or the host cell to a composition
CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC improving the production of PAL, its analog or another optically active
CC unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC are useful for treating a mammal having a disease, disorder or condition
CC selected from phenylketonuria, cancer, human immunodeficiency virus
CC infection and human cytomegalovirus infection. The present sequence
CC represents a R. graminis PAL polypeptide encoding cDNA

XX Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other;

Query Match 100.0%; Score 2419.6; DB 6; Length 2419;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTGCTCACTTAACCGGTGCTAGAGCTCGAGATGGCCCTTCTTGAATGGCTC 60
DB 1 CTCTGCTCACTTAACCGGTGCTAGAGCTCGAGATGGCCCTTCTTGAATGGCTC 60
QY 61 GCCACCAAGCTGCGCAACGGCTTTACCAACGGCTGCAAGCGCTTCCGACCAAGTGGCT 120
DB 61 GCCACCAAGCTGCGCAACGGCTTTACCAACGGCTGCAAGCGCTTCCGACCAAGTGGCT 120
QY 121 GCGGGGCGCCACTTGGGCTCGCGCGCAAGCGCGCTGATGGGCAAGCGCGCGACAG 180
DB 121 GCGGGGCGCCACTTGGGCTCGCGCGCAAGCGCGCTGATGGGCAAGCGCGCGACAG 180
QY 181 TCGAGCTCGAGATCGTGCAGAGAGCTCTCAAGCAAGCGCGCTGCAAGTGGCTC 240
DB 181 TCGAGCTCGAGATCGTGCAGAGAGCTCTCAAGCAAGCGCGCTGCAAGTGGCTC 240
QY 241 AGCGGCTACAGCTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
DB 241 AGCGGCTACAGCTCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 301 CGCGTCCAGAACGACGACGAGATCGCGGACCGCTGCAAGAGAGCTGCACTTCTCAG 360
DB 301 CGCGTCCAGAACGACGACGAGATCGCGGACCGCTGCAAGAGAGCTGCACTTCTCAG 360
QY 361 GCCAGCTTACAGACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 GCCAGCTTACAGACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 AGGATGAGATGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 480
DB 421 AGGATGAGATGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 480
QY 481 ACGCGGACGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 481 ACGCGGACGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 GAGGCTCGCTCGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 541 GAGGCTCGCTCGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 601 GTCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 601 GTCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 661 GTCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 GTCCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 GCGCGCTACACCGGCTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 721 GCGCGCTACACCGGCTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 781 ATCATGTTGCGCGGCGGCGCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 ATCATGTTGCGCGGCGGCGCATCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 AAGAGGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900

DB 841 AAGAGGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 901 CTGACAGATCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 901 CTGACAGATCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 961 GCGATGCTCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 GCGATGCTCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1021 CCGG 1080
DB 1021 CCGG 1080
QY 1081 GTTACGACGACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 GTTACGACGACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 1141 TACCTGCTCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 TACCTGCTCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 TACTGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1201 TACTGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1261 AACAGGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
DB 1261 AACAGGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
QY 1321 AAGACGAGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 AAGACGAGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
QY 1381 CTCAAGCT 1440
DB 1381 CTCAAGCT 1440
QY 1441 AACTATACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 AACTATACGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 CTTGCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
DB 1501 CTTGCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1561 TCGCT 1620
DB 1561 TCGCT 1620
QY 1621 CTGCT 1680
DB 1621 CTGCT 1680
QY 1681 TTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
DB 1681 TTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
QY 1741 CTGACGCTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
DB 1741 CTGACGCTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
QY 1801 ACGACGACGCT 1860
DB 1801 ACGACGACGCT 1860
QY 1861 GTGCTGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
DB 1861 GTGCTGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
QY 1921 TGAAGGTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980

```
Db 1921 TGAAGGTTGGCTGGCCGAGAAAGCATTCTCCCTACGCGCGAGGTGGCAACGCTTC 1980
2y 1991 TGGAGAGCGCGCTTTGGCAGCGCGCGCGAGCATCTCTCGCGCGCAGCGCGTC 2040
Db 1991 TGGAGAGCGCGCTTTGGCAGCGCGCGCGAGCATCTCTCGCGCGCAGCGCGTC 2040
2y 2041 CTGTAATCTGTTGGCGGAGAGAGCTCGGCGTGGAGCGCGCGCGAGCGTGTTC 2100
Db 2041 CTGTAATCTGTTGGCGGAGAGAGCTCGGCGTGGAGCGCGCGAGCGTGTTC 2100
2y 2101 GCGGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2101 GCGGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
2y 2161 GCGGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 GCGGTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
2y 2221 GACGCGCGCTCAACCCCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 GACGCGCGCTCAACCCCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
2y 2281 CCTCATACAGATGCGCTTACTCTCTCGCGCTCATACAGTCTCTAGTTCTTGAT 2340
Db 2281 CCTCATACAGATGCGCTTACTCTCTCGCGCTCATACAGTCTCTAGTTCTTGAT 2340
2y 2341 CCGCGGTCTCGGTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2341 CCGCGGTCTCGGTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
2y 2401 AGTCAAAAAAAAAAAAAA 2419
Db 2401 AGTCAAAAAAAAAAAAAA 2419

RESULT 2
AAd5812
ID AAd5812 standard; DNA; 2419 BP.
XX
AC AAd5812;
XX
DT 27-DEC-2002 (first entry)
DE Rhodotorula graminis PAL DNA #1.
XX
KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytostatic; virucide; gene; ds.
XX
OS Rhodotorula graminis.
XX
FH Key Location/Qualifiers
FT CDS 37..2199
FT /tag= a
FT /product= "PAL protein"
FT /EC number= "4.3.1.5"
FT /transl_except= (pos:493..495, aa:Xaa)
FT /note= "Xaa corresponds to Val or Ala"
XX
XX US2002102712-A1.
XX
XX 01-AUG-2002.
XX
XX 24-AUG-2001; 2001US-00939408.
XX
XX 24-JUL-2000; 2000US-00624693.
XX
XX 24-JUL-2001; 2001MO-US023270.
XX
XX (PCBU-) PCBU SERVICES INC.
XX
XX Yoshida RK, Kooester AB;
XX
XX WPI, 2002-690616/74.
DR
```

```
DR P-PSDB; AAE27938.
XX
PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT human immunodeficiency virus or human cytomegalovirus infection.
XX
PS Claim 13; Page 27-29; 74pp; English.
XX
CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC such proteins. PAL sequences are useful for producing L-phenylalanine.
CC They are useful for treating mammals having diseases, disorders or
CC conditions that would benefit from treatment with PAL proteins such as
CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC also used in gene therapy. The present sequence is R. graminis PAL DNA
XX
SQ Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other.
XX
Query Match 100.0%; Score 2418.6; DB 6; Length 2419;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 CTCCTGCTCACTTAACCGCTACAGTGCATGCGATGCGCCCTTCTTGACTGCTC 60
Db 1 CTCCTGCTCACTTAACCGCTACAGTGCATGCGATGCGCCCTTCTTGACTGCTC 60
2y 61 GCCACAGAGCTCGCCAAAGGCTTTACCAAGGCTGACAGCGCTCGACCAAGTGGCT 120
Db 61 GCCACAGAGCTCGCCAAAGGCTTTACCAAGGCTGACAGCGCTCGACCAAGTGGCT 120
2y 121 GCGGGCCCACTTGGCTCTTCGCGCGACAGCGCGGCTCGATGAGCGCGCGACAG 180
Db 121 GCGGGCCCACTTGGCTCTTCGCGCGACAGCGCGGCTCGATGAGCGCGCGACAG 180
2y 121 GCGGGCCCACTTGGCTCTTCGCGCGACAGCGCGGCTCGATGAGCGCGCGACAG 180
Db 121 GCGGGCCCACTTGGCTCTTCGCGCGACAGCGCGGCTCGATGAGCGCGCGACAG 180
2y 181 TCGAGCTGCGAGATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 181 TCGAGCTGCGAGATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
2y 241 AGCGGTACAGCTCACCGTCCGTGAGAGTTGTGCGCGCGCGCGCAAGGGGCGAGG 300
Db 241 AGCGGTACAGCTCACCGTCCGTGAGAGTTGTGCGCGCGCGCGCAAGGGGCGAGG 300
2y 301 CCGGTCCAGAGAGAGAGAGAGAGATCCGCGAGCGCTCGACAAAGAGCGTCTTCAAG 360
Db 301 CCGGTCCAGAGAGAGAGAGAGATCCGCGAGCGCTCGACAAAGAGCGTCTTCAAG 360
2y 361 GCGGTCCAGAGAGAGAGAGAGATCCGCGAGCGCTCGACAAAGAGCGTCTTCAAG 420
Db 361 GCGGTCCAGAGAGAGAGAGAGATCCGCGAGCGCTCGACAAAGAGCGTCTTCAAG 420
2y 421 AGGACTGAGAGATGAGTACAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 421 AGGACTGAGAGATGAGTACAGCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
2y 481 AGCGGAGAGTCCGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Db 481 AGCGGAGAGTCCGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
2y 541 GAGGTGTCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db 541 GAGGTGTCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
2y 601 GTCCGCTGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 GTCCGCTGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
2y 661 GTCCGCTGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 GTCCGCTGCTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
2y 721 GCGGCATACCGGTACCCCGAGCTCAAGGTTACGTTTGCACAGAGAGAGAGAGAG 780
Db 721 GCGGCATACCGGTACCCCGAGCTCAAGGTTACGTTTGCACAGAGAGAGAGAGAG 780
```


449. .880
/tag= e
/number= 2
881. .960
/tag= f
/number= 2
961. .1295
/tag= g
/number= 3
1296. .1364
/tag= h
/number= 3
1365. .1529
/tag= i
/number= 4
1530. .1586
/tag= j
/number= 4
1587. .1748
/tag= k
/number= 5
1749. .1821
/tag= l
/number= 5
1822. .1947
/tag= m
/number= 6
1948. .2007
/tag= n
/number= 6
2008. .2586
/tag= o
/number= 7
WO200208402-A2.
31-JAN-2002.
24-JUL-2001; 2001WO-US023270.
24-JUL-2000; 2000US-00624693.
(PCBU-) PCBU SERVICES INC.
Yoshida RK, Koetsura AB;
MPI: 2002-268973/31.
P-PsDB; ABB07690.
Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
treating mammal having disease or disorder from phenylketonuria, cancer,
human immunodeficiency virus infection and human cytomegalovirus
infection.
Claim 4; Fig 3, 4; 135p; English.
The invention relates to an isolated and purified yeast phenylalanine
ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
host cell comprising the polynucleotide is useful for the production of L
-PAL. By adding the polynucleotide or the host cell to a composition
comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
improving the production of PAL, its analog or another optically active
unnatural amino acid having PAL-like structure. The PAL polynucleotides
are useful for treating a mammal having a disease, disorder or condition
selected from phenylketonuria, cancer, human immunodeficiency virus
infection and human cytomegalovirus infection. The present sequence
represents a R. graminis PAL genomic DNA sequence
Sequence 2741 BP; 441 A; 1019 C; 756 G; 524 T; 0 U; 1 Other;
Query Match 75.5%; Score 1826.6; DB 6; Length 2741;
Best Local Similarity 84.3%; Pred. No. 1.5e-280;
Matches 2312; Conservative 2; Mismatches 1; Indels 426; Gaps 6;

QY 37 ATGGCCCTTCTTTGAGCTCGCTCGCCACGAGCTCGCCAAAGCTTTACAAAGGCTCG 96
DB 1 ATGCMCTTCTCTTGAATCTGCTCGCCACGCTCGCAAGGCTTTACAAAGGCTCG 60
QY 97 CAGCGCTTCGACCAAGTCTGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 156
DB 61 CAGCGCTTCGACCAAGTCTGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 120
QY 157 CTGATGAGCG 216
DB 121 CTGATGAGCG 180
QY 217 CCAACGACGAGCTGCTGAGCTCAGCGGCTACAGCTCAGCTCAGCTCAGCTCAGCTCAG 276
DB 181 CCAACGACGAGCTGCTGAGCTCAGCGGCTACAGCTCAGCTCAGCTCAGCTCAGCTCAG 240
QY 277 GCG 336
DB 241 GCG 300
QY 337 GACAAGAGCTGCACTTCTCAAGGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAG 395
DB 301 GACAAGAGCTGCACTTCTCAAGGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAG 360
QY 396 ----- 395
DB 361 GTGCTTCCGACGAGAGGCGGAAATCTCGGATGCGGAGCTGAAAGCTGACATC 420
QY 396 -----CGTTTGGTGGCTCGGCGCGACGAGAGCTGAG 429
DB 421 GCTTGAAGGCTCGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGG 480
QY 430 GATGAGTCAAGCTTCAGAAAGGCTGATGAGACGAGCTTGGCGGCTGAGCGCGAG 489
DB 481 GATGAGTCAAGCTTCAGAAAGGCTGATGAGACGAGCTTGGCGGCTGAGCGCGAG 540
QY 490 TCGGCTCGCTTCAAGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 549
DB 541 TCGGCTCGCTTCAAGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 600
QY 550 CGGCGCGCATGATCATCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 609
DB 601 CGGCGCGCATGATCATCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 660
QY 610 GTGCTCTTGAAGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAG 669
DB 661 GTGCTCTTGAAGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAG 720
QY 670 CGGCGCTCATCTCGGCGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 729
DB 721 CGGCGCTCATCTCGGCGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGG 780
QY 730 ACCGCTACCCCGAGCTGAGGTTGAGGTTTGGCAGAGGAAACCGAAGATCATGTTT 789
DB 781 ACCGCTACCCCGAGCTGAGGTTGAGGTTTGGCAGAGGAAACCGAAGATCATGTTT 840
QY 790 GCGCGCGAGGCTGAGGTTTGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 828
DB 841 GCGCGCGAGGCTGAGGTTTGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 900
QY 829 -----G 829
DB 901 GAGAGAGCTGTGAGAGTCTCCAGTTTCTGATGCTGATGCTGATGCTGATGCTGATG 960
QY 830 TCTTGGCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 889
DB 961 TCTTGGCGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
QY 890 CGACCTGAGTGTGACAGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 949
DB 1021 CGACCTGAGTGTGACAGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080

QY 950 TCACGATGAGGACCATGATGATGAGGACGAGGCTGATGAGGCTTCAATCCAGACGCT 1009
 DB 1081 TCACGATGAGGACCATGATGATGAGGACGAGGCTGATGAGGCTTCAATCCAGACGCT 1140
 QY 1010 GCCGCGGACACCCGCGACAGGTGAGGTGCGCGCAACATCCGCAAGCTCTTCCGAGCT 1069
 DB 1141 GCCGCGGACACCCGCGACAGGTGAGGTGCGCGCAACATCCGCAAGCTCTTCCGAGCT 1200
 QY 1070 GGTGCTTGGCGGCTTGAAGCAAGAGAGAGTCAAGTCAAGAGCAAGAGGCAATTTCTC 1129
 DB 1201 GGTGCTTGGCGGCTTGAAGCAAGAGAGAGTCAAGTCAAGAGCAAGAGGCAATTTCTC 1260
 QY 1130 GCCAGAGCCGCTACCCGCTCCGCAAGTCCCT-----CAGTTCTCGGCGGCTCG 1161
 DB 1261 GCCAGAGCCGCTACCCGCTCCGCAAGTCCCTCAAGTTCTCTCTCTCTCTCTCTCTCC 1320
 QY 1162 -----CAGTTCTCGGCGGCTCG 1180
 DB 1321 CTCGCTCCGACCGGCGGCTGAGACTTACGTTTTCGATTCAGTTCTCTCGGCGGCTCG 1380
 QY 1181 TGGAGACATGATGACACGCTTACTGATCTCTGCTGAGAGCAACAACAGAGAGAGCA 1240
 DB 1381 TGGAGACATGATGACACGCTTACTGATCTCTGCTGAGAGCAACAACAGAGAGAGCA 1440
 QY 1241 ACCCGCTCTGACGCTGAGAAACAAGCAACCGCGCAAGCGGCACTTCCAGGAGCTCG 1300
 DB 1441 ACCCGCTCTGACGCTGAGAAACAAGCAACCGCGCAAGCGGCACTTCCAGGAGCTCG 1500
 QY 1301 CTGCTCGATTTGATGAGAGAAC----- 1325
 DB 1501 CTGCTCGATTTGATGAGAGAACCAAGTGTCTCTGCTGCTTCTGATCTCCGATCT 1560
 QY 1326 -----CAAGCTGCACTGCGCTTCACTGAGAGCTCAACTTCA 1363
 DB 1561 TGTGTAATGTTCTTCTCTGAGAGCTGCACTGCGCTTCACTGAGAGCTCAACTTCA 1620
 QY 1364 CGCAGTGCACCGAGTGTCTCAACGCTGCACTGAAACCGCGCTCTGCTGCTGCTG 1423
 DB 1621 CGCAGTGCACCGAGTGTCTCAACGCTGCACTGAAACCGCGCTCTGCTGCTGCTG 1680
 QY 1424 CGCAGTGCACCGCTGCTCACTATCAAGGAGGCTTGAATTCATCGATCGATGCTTACG 1483
 DB 1681 CGCAGTGCACCGCTGCTCACTATCAAGGAGGCTTGAATTCATCGATCGATGCTTACG 1740
 QY 1484 CTTCGG----- 1489
 DB 1741 CTTCGGAGGTGAGCCGTGAGCTTCTCGCGCTGCTGCTCCCTTCAAGGACACCGAGG 1800
 QY 1490 -----AGCTGCGGCACTTTCGCAACCGGTCCTACTTCTGCTCAG 1530
 DB 1801 TGAATTTCTTCCCTCTGATGCTGAGGACCTTCTGCGCTGCGACTGATGAGGCTCAG 1860
 QY 1531 CCGCAGATGAGGACACAGGCGCTCACTGCTGCTCACTCTGCGCGCGCGAGCT 1590
 DB 1861 CCGCAGATGAGGACACAGGCGCTCACTGCTGCTCACTCTGCGCGCGCGAGCT 1920
 QY 1591 GCCGAGGCGCAAGAGCTCTTCTCTC----- 1617
 DB 1921 GCCGAGGCGCAAGAGCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
 QY 1618 -----CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1650
 DB 1981 AGCGACTGATGCGGATCTGAGACGCTTCTGCGCTGCGACTGATGAGGCTCAG 2040
 QY 1651 GCCGTCGACCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1710
 DB 2041 GCCGTCGACCTGCGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2100
 QY 1711 ACTCTCTCTGACGACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1770
 DB 2101 ACTCTCTCTGACGACCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 2160
 QY 1771 AAGAAGGCGCTCAACAAGGCTTCTGAGAGACAGACGATCGACTGAGCGGCTGAG 1830

DB 2161 AAGAAGGCGCTCAACAAGGCTTCTGAGAGACAGACGATGACCTTCTGAGCGGCTGAG 2220
 QY 1831 CACGAGGCTTCTGATGACGACGAGGACGCTGCTGAGCTCTCTGCTGCTGCTCT 1890
 DB 2221 CACGAGGCTTCTGATGACGACGAGGACGCTGCTGAGCTCTCTGCTGCTGCTCT 2280
 QY 1891 GCCAAGCTGACCTTACTGCTGCAACGCTGGAAGGTTGCTGAGCGGAGGAGCTC 1950
 DB 2281 GCCAAGCTGACCTTACTGCTGCAACGCTGGAAGGTTGCTGAGCGGAGGAGCTC 2340
 QY 1951 TCGCTCAGCGGAGGAGTGCAGACCGCTTCTGAGAGCGCGCTTCTGAGCGGCGG 2010
 DB 2341 TCGCTCAGCGGAGGAGTGCAGACCGCTTCTGAGAGCGCGCTTCTGAGCGGCGG 2400
 QY 2011 CACGATACCTTCTGCGCGGACGCGGCTCTGATCTGCTGCTGCTGCTGCTGCTGCTG 2070
 DB 2401 CACGATACCTTCTGCGCGGACGCGGCTCTGATCTGCTGCTGCTGCTGCTGCTGCTG 2460
 QY 2071 GTGAGGCGGCGCGGCGGAGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2130
 DB 2461 GTGAGGCGGCGCGGCGGAGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520
 QY 2131 GTCTGCGCATCTTACGAGGACATCAAGAGAGCGCGCATCAACAGCTCTGCTGAGATG 2190
 DB 2521 GTCTGCGCATCTTACGAGGACATCAAGAGAGCGCGCATCAACAGCTCTGCTGAGATG 2580
 QY 2191 CTGCGCTGAGGCGCGGAGAGGCTCTGCTGAGAGCGCGCTCAACCGAAGACGCTTT 2250
 DB 2581 CTGCGCTGAGGCGCGGAGAGGCTCTGCTGAGAGCGCGCTCAACCGAAGACGCTTT 2640
 QY 2251 CGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2310
 DB 2641 CGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2700
 QY 2311 CCGTCATCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2351
 DB 2701 CCGTCATCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2741

RESULT 4
 ADD45823
 ID ADD45823 standard; DNA; 2741 BP.
 XX
 AC ADD45823;
 DT 27-DEC-2002 (first entry)
 XX
 DE Rhodocorylia graminis PAL genomic DNA.
 XX
 KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
 KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
 KW human cytomegalovirus infection; cytostatic; virocid; gene; ds.
 OS Rhodocorylia graminis.
 XX
 FH Key
 FT CDS
 FT 1..2589 Location/Qualifiers
 FT /*tag= a
 FT /product= "PAL protein"
 FT /EC_number= "4.3.1.5"
 FT 1..361
 FT /*tag= b
 FT 362..448
 FT /*tag= c
 FT 449..880
 FT /*tag= d
 FT 881..960
 FT /*tag= e
 FT 961..1285
 FT /*tag= f
 FT 1286..1364
 FT /*tag= g
 FT intron

FT exon 1355..1529
FT /tag= h
FT intron 1530..1586
FT /tag= 1
FT exon 1587..1748
FT /tag= j
FT intron 1749..1821
FT /tag= k
FT exon 1822..1947
FT /tag= l
FT intron 1948..2007
FT /tag= m
FT exon 2008..2589
FT /tag= n
FT terminator 2587..2589
FT /tag= o

US2002102712-A1.

01-AUG-2002.

24-AUG-2001; 2001US-00939408.

24-JUL-2000; 2000US-00624693.

24-JUL-2001; 2001WO-US023270.

(PCBU-) PCBU SERVICES INC.

Yoshida RK, Kootstra AB;

WPI; 2002-690616/74.

Novel isolated and purified Rhodotorula phenylalanine ammonia lyase polypeptide, useful for treating a mammal having phenylketonuria, cancer, human immunodeficiency virus or human cytomegalovirus infection.

Claim 4; Page 50-53; 74p; English.

The present invention relates to yeast (e.g. Rhodotorula) phenylalanine ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding such proteins. PAL sequences are useful for producing L-phenylalanine. They are useful for treating mammals having diseases, disorders or conditions that would benefit from treatment with PAL proteins such as phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or human cytomegalovirus (HCMV) infection. Sequences of the invention are also used in gene therapy. The present sequence is R. graminis PAL genomic DNA

Sequence 2741 BP; 441 A; 1019 C; 756 G; 524 T; 0 U; 1 Other;

Query Match 75.5%; Score 1826.6; DB 6; Length 2741;

Best Local Similarity 84.3%; Pred. No. 1.5e-280;
Matches 2312; Conservative 2; Mismatches 1; Indels 426; Gaps 6;

QY 37 ATGGCCCTTCTTGACTCGCTCGCACACACGCTCGCAACGGCTTTACCAAGGCTCG 96
DB 1 ATGGCCCTTCTTGACTCGCTCGCACACGCTCGCAACGGCTTTACCAAGGCTCG 60
QY 97 CAGCGCGCTCCCAACAAGTCGCTGCGGCGCCCACTTCGGCTTCGCGGACGCCGCG 156
DB 61 CAGCGCGCTCCCAACAAGTCGCTGCGGCGCCCACTTCGGCTTCGCGGACGCCGCG 120
QY 157 CTGATGCGCAGCGCGCAGCAGTGCAGTGCAGATGTCAGAGACTCTCAGCGAC 216
DB 121 CTGATGCGCAGCGCGCAGCAGTGCAGTGCAGATGTCAGAGACTCTCAGCGAC 180
QY 217 CCCACGACGAGCTGTCAGTCACTCAGCGGCTACAGCTTCACCTTCAGAGCTTCGCG 276
DB 181 CCCACGACGAGCTGTCAGTCACTCAGCGGCTACAGCTTCACCTTCAGAGCTTCGCG 240
QY 277 GCGCGCGCAGGCGGCGCAGGCTCGGCTCCAGAAAGCAGCAGAGATCCGCGCAAGCTC 336
DB 241 GCGCGCGCAGGCGGCGCAGGCTCGGCTCCAGAAAGCAGCAGAGATCCGCGCAAGCTC 300

QY 337 GACAAGAGCTCGACTCTCTCAAGGCCAGCTTCAGAACTCGGTCTACCGAGTCAACAC 395
DB 301 GACAAGAGCTCGACTCTCTCTCAAGGCCAGCTTCAGAACTCGGTCTACCGAGTCAACAC 360
QY 396 ----- 395
DB 361 GTGCGTCCGAGACGAGAGCGGAAAATCTCGGAGTCCGACGCTGAAACGCTGACATTC 420
QY 396 -----GGTTTCGTGCTCGCGCCGACACGAGAGCTGAG 429
DB 421 GCTTGAACGGCTCGCGGCTCTTGCAAGGTTTCGTGCTCGCGCCGACACGAGAGCTGAG 480
QY 430 GATCAGTCAAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCGGACG 540
DB 481 GATCAGTCAAGCTTCAGAAAGCGCTCATGAGACACAGCTTCGCGCGTGAACGCGGACG 540
QY 490 TCCGCTCTGCTTCAAGCTTCGAGCGGCTCGAAGAACAGCTTCGCTCGAGGTCGTC 549
DB 541 TCCGCTCTGCTTCAAGCTTCGAGCGGCTCGAAGAACAGCTTCGCTCGAGGTCGTC 600
QY 550 CCGCGCGCATGCTCATCGCGTCACTCGCTCAGCGGTCGACCTCGGCGCTCGGCTTC 609
DB 601 CCGCGCGCATGCTCATCGCGTCACTCGCTCAGCGGTCGACCTCGGCGCTCGGCTTC 660
QY 610 GTCTCTTGAAGCGCTCACCACCTTTGAACACCGCATCACGCCCATCTGCTCCCTTC 669
DB 661 GTCTCTTGAAGCGCTCACCACCTTTGAACACCGCATCACGCCCATCTGCTCCCTTC 720
QY 670 CCGCGCTTCATCTCGGCTCGGCGGAGACTCAGCGGCTCTGACATCGCGCGCATC 729
DB 721 CCGCGCTTCATCTCGGCTCGGCGGAGACTCAGCGGCTCTGACATCGCGCGCATC 780
QY 730 ACCGCTCAACCCCGACGTCACAGTTTCAAGTTTGAACGAGGAAACGAGAAATATGTTT 789
DB 761 ACCGCTCAACCCCGACGTCACAGTTTCAAGTTTGAACGAGGAAACGAGAAATATGTTT 840
QY 790 GCGCGCGAGGCATCTCGCTCTTTGCTCTGAGGCACTC----- 828
DB 841 GCGCGCGAGGCATCTCGCTCTTTGCTCTGAGGCACTC-----G 900
QY 829 -----G 829
DB 901 GCAGTGAAGCTTTCAGAGTCTCCCAAGTTTCTGCTACGCTTCCTTGTTCATGCGATTGAG 960
QY 830 TCTTGGGCGGAGAGGAGGCTCGGCTGCTGTAACGAGAGCGGCTTCGCGCTCATGAG 889
DB 961 TCTTGGGCGGAGAGGAGGCTCGGCTGCTGTAACGAGAGCGGCTTCGCGCTCATGAG 1020
QY 890 CGACCTTCAGTCTGACAGTCTGCAATGCTGCTGCTGCTGCTGCAAGGCTTGAAGGCTC 949
DB 1021 CGACCTTCAGTCTGACAGTCTGCAATGCTGCTGCTGCTGCTGCAAGGCTTGAAGGCTC 1080
QY 950 TCACGTTGAAGGCTTGAAGTCTGCGGCGCAGAGGCTGCTGCGCGCTTCATCAGAGCTCT 1009
DB 1081 TCACGTTGAAGGCTTGAAGTCTGCGGCGCAGAGGCTGCTGCGCGCTTCATCAGAGCTCT 1140
QY 1010 GCGCGCGCAGCCCGCGCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1069
DB 1141 GCGCGCGCAGCCCGCGCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1200
QY 1070 CGTCTGTTGCGGTTGAGACGAGAGGAGGTCAGAGTCAAGTCAAGTCAAGTCAAGTCAAG 1129
DB 1201 CGTCTGTTGCGGTTGAGACGAGAGGAGGTCAGAGTCAAGTCAAGTCAAGTCAAGTCAAG 1260
QY 1130 GCGAGACCGCTACCGCGCTTCGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1161
DB 1261 GCGAGACCGCTACCGCGCTTCGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1320
QY 1162 -----CAGTTCTCGCGCGCTCG 1180
DB 1321 CTCCTGCGACCGCGCGCTTCGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1380

QY 1181 TGAAGACATGATGACAGCCCTACTGACTCTCTCGCTCGAAGAACACAGAGACGACA 1240
Db 1381 TGAAGACATGATGACAGCCCTACTGACTCTCTCGCTCGAAGAACACAGAGACGACA 1440
QY 1241 ACCCGCTCCCGAGAGTGGAGAACAGAACACCGCGGACAGGCGGCACTTCCAGGGCTCGG 1300
Db 1441 ACCCGCTCCCGAGAGTGGAGAACAGAACACCGCGGACAGGCGGCACTTCCAGGGCTCGG 1500
QY 1301 CTGCTCGATTTGATGAGAGAGC----- 1325
Db 1501 CTGCTCGATTTGATGAGAGAGACAGAGTGGCTCTCTCGCTCGCTTCTGACTCCGATCT 1560
QY 1326 -----CAGGCTGCACTGCGCTCATGCGCAAGCTCAACTTCA 1363
Db 1561 TGTGCTGAATGTTCTTCTCTGAGGCTCGACCTCGCCCTCATCGGCAAGCTCAACTTCA 1620
QY 1364 CGCAGTGCACCGAGTTGCTCAAGCTGCAATGAACCGGCGCTGCGCTTCTGCGCTG 1423
Db 1621 CGCAGTGCACCGAGTTGCTCAAGCTGCAATGAACCGGCGCTGCGCTTCTGCGCTG 1680
QY 1424 CCGAGGACCCGCTGCTCAACTATCAAGGAGGCTTGAATTCACATCGCTGCTTACG 1483
Db 1681 CCGAGGACCCGCTGCTCAACTATCAAGGAGGCTTGAATTCACATCGCTGCTTACG 1740
QY 1484 CTTGCG----- 1489
Db 1741 CTTGCGAGGTGAGCCGCTGCAAGCTTCTCGCGCTCGCTGCTGCTGCGCAATGATCCGACG 1800
QY 1490 -----AGCTCGGCACTTGGCAACCGGAGTCACTGACTGCTGCTGAG 1530
Db 1801 TGACTTCCTTCCCTCTGTAGTTCGCGCACCTTGGCAACCGGAGTCACTGACTGCTGAG 1860
QY 1531 CCGCGAGATGAGGCAACCGGCGCTGCAACTGCTGCTCATCTCGCGCGCGCACT 1590
Db 1861 CCGCGAGATGAGGCAACCGGCGCTGCAACTGCTGCTCATCTCGCGCGCGCACT 1920
QY 1591 GCGAGAGCCCAAGAGTCTCTCTC----- 1617
Db 1921 GCGAGAGCCCAAGAGTCTCTCTCTGCGGCTTGTGTGCAATGATCCGACGCAAT 1980
QY 1618 -----CTTCTCGCTTGCACCTGTACTGACAGCTTCAG 1650
Db 1981 AGCGACTGACTCGCGATCTGTAGAGCACTTCTCGCTTGCACCTGTAAG 2040
QY 1651 GCGCTGCACTCCGCGCATGAGTGCATTTCAAGAGAGATTGCAAGCTTCTTCCG 1710
Db 2041 GCGCTGCACTCCGCGCATGAGTGCATTTCAAGAGAGATTGCAAGCTTCTTCCG 2100
QY 1711 ACTCTCTCAAGACACCTCGGCACTGCTGCAAGTCAAGCACTTGGCTCGAGTTC 1770
Db 2101 ACTCTCTCAAGACACCTCGGCACTGCTGCAAGTCAAGCACTTGGCTCGAGTTC 2160
QY 1771 AAGAGGCGCTCAACAGCGTCTCGAGAGAGAGTCAAGCACTTGGCGCGCTG 1830
Db 2161 AAGAGGCGCGCTCAACAGCGTCTCGAGAGAGAGTCAAGCACTTGGCGCGCTG 2220
QY 1831 CACGAGCGCTTGTAGAGGAGCGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1890
Db 2221 CACGAGCGCTTGTAGAGGAGCGGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280
QY 1891 GCGAAGCTCACCTTACTGCGGTCAAGCGTGAAGAGTTCCTGCGCGGAGAGGCAATC 1950
Db 2281 GCGAAGCTCACCTTACTGCGGTCAAGCGTGAAGAGTTCCTGCGCGGAGAGGCAATC 2340
QY 1951 TCGCTCAGCGGAGGAGTCCGCAACCGCTTCTGGCAAGCGGCTTTCGAGCGCGCG 2010
Db 2341 TCGCTCAGCGGAGGAGTCCGCAACCGCTTCTGGCAAGCGGCTTTCGAGCGCGCG 2400
QY 2011 CACGATACCTCTGCGCGGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2070
Db 2401 CACGATACCTCTGCGCGGCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2460
QY 2071 GTGCAAGGCGCGCGCGGCAAGCTGTTGTGTGCGCGTTCAGAGAGAGAGTCCGAGCAAC 2130

Db 2461 GTGCAAGGCGCGCGCGGCAAGCTGTTGTGTGCGCGTTCAGAGAGAGAGTCCGAGCAAC 2520
QY 2131 GTCTGCGCATTTACAGGCGCATCAAGGAGCGCGCATCAACAGCTCTGCTGAGATG 2190
Db 2521 GTCTGCGCATTTACAGGCGCATCAAGGAGCGCGCATCAACAGCTCTGCTGAGATG 2580
QY 2191 CTCGCGTAAAGCGCGAGAGCGCTCGCTTAAGCGCGCGCTCAACCGCAAGCAAGCTTTT 2250
Db 2581 CTCGCGTAAAGCGCGAGAGCGCTCGCTTAAGCGCGCGCTCAACCGCAAGCAAGCTTTT 2640
QY 2251 CGACGTGCTGTGTCGCAAGAGCACTTCTCTCATACAGATGCTGCTTACTCTG 2310
Db 2641 CGACGTGCTGTGTCGCAAGAGCACTTCTCTCATACAGATGCTTACTCTGCTG 2700
QY 2311 CCGTATACAGTCTCTGCTTCTTCTGATCCGCGTCTCT 2351
Db 2701 CCGTATACAGTCTCTGCTTCTTCTGATCCGCGTCTCT 2741

RESULT 5
AAd5824
ID AAd5824 standard; DNA; 2163 BP.
XX
AC AAd5824;
XX
DT 27-DEC-2002 (first entry)
XX
DE Yeast phenylalanine ammonia lyase DNA #2.
XX
KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
XX human cytomegalovirus infection; cytosstatic; virucide; ds.
OS Rhodocycla graminis.

Key
FH Location/Qualifiers
FT 1..2163
FT /*tag= a
FT /product= "PAL protein"
FT /EC number= "4.3.1.5"
FT /transl_except= (pos:13..15, aa:Xaa)
FT /note= "Xaa corresponds to Val, Leu, Phe"
FT /transl_except= (pos:34..36, aa:Xaa)
FT /note= "Xaa corresponds to Val, Leu, Phe"
FT /transl_except= (pos:46..48, aa:Xaa)
FT /note= "Xaa corresponds to Val, Leu, Phe"
FT /transl_except= (pos:49..51, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:55..57, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:58..60, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Leu or an in-frame stop codon"
FT /transl_except= (pos:73..75, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:79..81, aa:Xaa)
FT /note= "Xaa corresponds to Pro, Ser"
FT /transl_except= (pos:82..84, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro"
FT /transl_except= (pos:100..102, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:106..108, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:115..117, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT /transl_except= (pos:118..120, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Gly, Trp"
FT /transl_except= (pos:142..144, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Thr, Met, Glu, Ala, Val, Gln, Pro, Leu"
FT /transl_except= (pos:160..162, aa:Xaa)
FT /note= "Xaa corresponds to Val"

FT /transl_except= (pos:166. .168, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Glu, Gln"
FT /transl_except= (pos:193. .195, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp, Val"
FT /transl_except= (pos:196. .198, aa:Leu)
FT /note= "Xaa corresponds to Ile, Val, Leu"
FT /transl_except= (pos:226. .228, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:259. .261, Pro, Ser)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:277. .279, aa:Xaa)
FT /note= "Xaa corresponds to Asp, Ala"
FT /transl_except= (pos:304. .306, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Asn"
FT /transl_except= (pos:307. .309, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:325. .327, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:334. .336, aa:Xaa)
FT /note= "Xaa corresponds to Asp, His, Tyr"
FT /transl_except= (pos:340. .342, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:349. .351, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:442. .444, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:448. .450, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:457. .459, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Phe"
FT /transl_except= (pos:460. .462, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp"
FT /transl_except= (pos:469. .471, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Gly, Arg"
FT /transl_except= (pos:475. .477, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:547. .549, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:667. .669, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:673. .675, aa:Xaa)
FT /note= "Xaa corresponds to Ser"
FT /transl_except= (pos:709. .711, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Thr, Ile, Gly, Ala, Val,
Cys, Phe"
FT /transl_except= (pos:715. .717, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:721. .723, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:724. .726, aa:Xaa)
FT /note= "Xaa corresponds to Val, Leu, Phe"
FT /transl_except= (pos:736. .738, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Asn, Thr"
FT /transl_except= (pos:751. .753, aa:Xaa)
FT /note= "Xaa corresponds to Tyr, Ser, Phe"
FT /transl_except= (pos:775. .777, aa:Xaa)
FT /note= "Xaa corresponds to Leu, Phe"
FT /transl_except= (pos:913. .915, aa:Xaa)
FT /note= "Xaa corresponds to Leu"
FT /transl_except= (pos:955. .957, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1036. .1038, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Arg, Thr, Gln, Pro, Trp,
Ser or a stop codon"
FT /transl_except= (pos:1231. .1233, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Thr, Met"
FT /transl_except= (pos:1261. .1263, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1369. .1371, aa:Xaa)
FT /note= "Xaa corresponds to Leu"
FT /transl_except= (pos:1372. .1374, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1396. .1398, aa:Xaa)

FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1459. .1461, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:1477. .1479, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:1498. .1500, aa:Arg)
FT /transl_except= (pos:1552. .1554, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:1606. .1608, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:1666. .1668, aa:Xaa)
FT /note= "Xaa corresponds to Met, Ile, Val, Leu"
FT /transl_except= (pos:1669. .1671, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Leu"
FT /transl_except= (pos:1672. .1674, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Pro"
FT /transl_except= (pos:1675. .1677, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1681. .1683, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Leu, Phe"
FT /transl_except= (pos:1684. .1686, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Glu, Gln"
FT /transl_except= (pos:1699. .1701, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1714. .1716, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Tyr"
FT /transl_except= (pos:1723. .1725, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Ala, Val"
FT /transl_except= (pos:1741. .1743, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1747. .1749, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Asp, Tyr"
FT /transl_except= (pos:1849. .1851, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1852. .1854, aa:Xaa)
FT /note= "Xaa corresponds to Pro, Ser"
FT /transl_except= (pos:1927. .1929, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
stop codon"
FT /transl_except= (pos:1939. .1941, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:1948. .1950, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT /transl_except= (pos:1978. .1980, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:2122. .2124, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
Tyr, Cys"
FT /transl_except= (pos:2128. .2130, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Arg, Cys"
XX US2002102712-A1.
XX 01-AUG-2002.
XX 24-AUG-2001; 2001US-00939408.
XX 24-JUL-2000; 2000US-00624693.
XX 24-JUL-2001; 2001WO-US023270.
XX (PCBU-) PCBU SERVICES INC.
XX Yoshida RK, Koestera AB,
XX WPI; 2002-690616/74.
XX P-PsDB; AAE27943.
XX Novel isolated and purified Rhodocorula phenylalanine ammonia lyase
XX polypeptide, useful for treating a mammal having phenylketonuria, cancer,
XX human immunodeficiency virus or human cytomegalovirus infection.
XX Claim 7; Page 53-56; 74pp; English.
PS

CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC such proteins. PAL sequences are useful for producing L-phenylalanine.

Query Match 70.4%; Score 1704; DB 6; Length 2163;
Best Local Similarity 84.0%; Pred. No. 3,9e-261;
Matches 1817; Conservative 109; Mismatches 236; Indels 0; Gaps 0;

```
QY 37 ATGGCCCTCTCTTGGAGTCTGCTGCGACCAAGCGTGGTAAACCAAGGCTCG 96
Db 1 ATGGCCCCCTCCCTCGACTGATCGGACTCTGATCGGACGGBTCDCAAAGGCTG 60
QY 97 CAGCGGCTCCGACCAAGTGTGGGCGGCCCACTTCGCTTCGCGCGACGCGCGC 156
Db 61 CAGCGGCTCCGACCAAGTGTGGGCGGCCCACTTCGCTTCGCGCGACGCGCGC 120
QY 157 CTCGATGACCAAGCGCGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 216
Db 121 CTCGCTCTCCCGCCACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 180
QY 217 CCCACCGACAGCTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 276
Db 181 CCCACCGACAGCTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 240
QY 277 GCGCGCGGACAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 336
Db 241 GCGCGCGGACAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 337 GACAGAGCGTGAATCTCTCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
Db 301 GACAGAGCGTGAATCTCTCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 397 GGTTCGTTGGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 456
Db 361 GGTTCGTTGGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 420
QY 457 ATGAGACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 516
Db 421 ATGAGACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 480
QY 517 GGTTCGTTGGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 576
Db 481 GGTTCGTTGGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 540
QY 577 TGGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
Db 541 TGGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
QY 637 TTGAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 696
Db 601 TTGAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 660
QY 697 CTCGACCGGCTCTGTAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 756
Db 661 CTCGACCGGCTCTGTAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
QY 757 GTTTTGAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 816
Db 721 GTTTTGAACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 817 CTCGAGGAGTGTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 876
Db 781 CTCGAGGAGTGTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 877 TCCGCTGATGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 936
Db 841 TCCGCTGATGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
QY 937 GCGCTTGAAGGCTCTCAAGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 996
Db 901 GCGCTTGAAGGCTCTCAAGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 960
QY 997 ATTCAGAGCTGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1056
```

```
Db 961 CTCACAGAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
QY 1057 CTCCTTCCGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1116
Db 1021 CTCCTGAGGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
QY 1117 GAGGCGATTTCTGCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
Db 1081 GAGGCGATTTCTGCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
QY 1177 CTCGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1236
Db 1141 CTCGATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1200
QY 1237 GACACCGGCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1296
Db 1201 GACACCGGCTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1260
QY 1297 TCGGCTGCTGATTTGATGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1356
Db 1261 TCGGCTGCTGATTTGATGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
QY 1357 AACCTTCAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1416
Db 1321 AACCTTCAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1380
QY 1417 CTCGCTGCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
Db 1381 CTCGCTGCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
QY 1477 GCTTACGCTTGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1536
Db 1441 GCTTACGCTTGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1500
QY 1537 GAGATGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1596
Db 1501 GAGATGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
QY 1597 GCGACAGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1656
Db 1561 GCGACAGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
QY 1657 GACCTCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1716
Db 1621 GACCTCGGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1680
QY 1717 CTCGAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1776
Db 1681 CTCGAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740
QY 1777 GCGCTTCAAGAGGCTCTGAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1836
Db 1741 GCGCTTCAAGAGGCTCTGAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
QY 1837 GCGCTTCTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1896
Db 1801 GCGCTTCTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
QY 1897 GTCACCTTATGCGGCTCTGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1956
Db 1861 GTCACCTTATGCGGCTCTGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
QY 1957 ACAGCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2016
Db 1921 ACAGCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1980
QY 2017 TACCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2076
Db 1981 TACCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
QY 2077 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2136
```

2041 GCCCGCCGCGGCGACGTCTTCTTCGCGACAGAGGTGACGATCGGACCAACGTCTCC 2100
2y 2137 CGCATCTACGAGGCGCATACAGAGCGCCGATACACCACTCTCTGTAAGATGCTCGG 2196
2b 2101 CGCATCTACGAGGCGCATACAGAGCGCCGATACACCACTCTCTGTAAGATGCTCGD 2160
2y 2197 TA 2198
2b 2161 TA 2162
RESULT 6
ABA95244
ID ABA95244 standard; DNA; 2475 BP.
XX ABA95244;
AC 29-AUG-2003 (revised)
XX 07-AUG-2003 (revised)
DT 10-JUN-2002 (first entry)
DT Rhodotorula PAL consensus polynucleotide sequence.
DE Rhodotorula PAL consensus polynucleotide sequence.
XX PAL; yeast; phenylalanine ammonia lyase; cinamic acid; cinamate;
XX cytosolic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
KM consensus; ds.
XX Rhodotorula graminis.
XX Rhodotorula mucilaginosa.
XX Rhodotorula mucilaginosa.
XX Rhodospiridium toruloides.
XX
FH Key Location/Qualifiers
FT misc_feature 1..2475
FT /tag= "a
FT /note= "n= A or C or G or T; "n" indicates no consensus
FT at that position"
PN W0200208402-A2.
XX 31-JAN-2002.
PD 24-JUL-2001; 2001MO-US023270.
XX 24-JUL-2000; 2000US-00624693.
PR (PCBU-) PCBU SERVICES INC.
XX (PCBU-) PCBU SERVICES INC.
XX Yoshida RK, Koestera AB,
PI P-PSDB; ABB07694.
DR WPI; 2002-268973/31.
PT phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT treating mammal having disease or disorder from phenylketonuria, cancer,
PT human immunodeficiency virus infection and human cytomegalovirus
PT infection.
XX
PS Claim 5; Fig 2; 135pp; English.
XX The invention relates to an isolated and purified yeast phenylalanine
XX ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
XX host cell comprising the polynucleotide is useful for the production of L
XX -PAL, by adding the polynucleotide or the host cell to a composition
XX comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
XX improving the production of PAL, its analog or another optically active
XX unnatural amino acid having PAL-like structure. The PAL polynucleotides
XX are useful for treating a mammal having a disease, disorder or condition
XX selected from phenylketonuria, cancer, human immunodeficiency virus
XX infection and human cytomegalovirus infection. The present sequence
XX represents a consensus polynucleotide sequence of the Rhodotorula PAL.
XX (Updated on 07-AUG-2003 to correct OS field.) (Updated on 29-AUG-2003 to
XX standardise OS field)

XX Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;
SQ
Query Match 57.3%; Score 1629; DB 6; Length 2475;
Best Local Similarity 78.0%; Pred. No. 2.9e-249;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;
Qy 37 ATGAGCCCTTCCTTGAGTACGCTCGCCACCAAGCTCGCCAAAGCTTTTACCAAGCGCTCG 96
Db 1 ATGAGCCCTTCCTTGAGTACGCTCGCCACCAAGCTCGCCAAAGCTTTTACCAAGCGCTCG 60
Qy 97 CAGCGCGCTTCGACCAAGTGGCTGGGGGCCCACTTCGGCTCTCCGCCGACAGC--CCG 154
Db 61 CAGCGCGCTTCGACCAAGTGGCTGGGGGCCCACTTCGGCTCTCCGCCGACAGC--CCG 120
Qy 155 GCGTCGATGGGCAAGCGCGGCGACCAAGTCGAGCTCGAGATCGTGCGAGAGCTCCTCAGCG 214
Db 121 GCTGCTCTCTCCGACCAAGTGGCTGGGGGCCCACTTCGGCTCTCCGCCGACAGC--CCG 180
Qy 215 ACCCGACCGAC--GACGTGCTGAGCTCAGCGGTAACAGCTCAGCTTCGCTGACGTTG 271
Db 181 ACCCGACCGAC--GACGTGCTGAGCTCAGCGGTAACAGCTCAGCTTCGCTGACGTTG 240
Qy 272 TCGGCG 330
Db 241 TCGGCG 300
Qy 331 CGCGTCGACCAAGCGCTCGACCTTCCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
Db 301 AAGATCGACCAAGCGCTCGACCTTCCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Qy 391 ACCAGCGGTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 450
Db 361 ACCAGCGGTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Qy 451 GCGCTCATCGAGCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510
Db 421 GCGCTCATCGAGCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
Qy 511 GAGCGCGCGCTCGAGCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
Db 481 GAGCGCGCGCTCGAGCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
Qy 571 GTCAAGCTGCTCAGCGCTGCG 630
Db 541 GTCAAGCTGCTCAGCGCTGCG 600
Qy 631 AACTCTTGAACCAAGCGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
Db 601 AACTCTTGAACCAAGCGCTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Qy 691 GAGCGCGCTCAGCGCTCAGCG 750
Db 661 GAGCGCGCTCAGCGCTCAGCG 720
Qy 751 GTTCAAGTTTTCAG 810
Db 721 GTTCAAGTTTTCAG 780
Qy 811 TTGCTCTGAG 870
Db 781 TTGCTCTGAG 840
Qy 871 GCGCTCTGCGCTCGATGCG 930
Db 841 GCGCTCTGCGCTCGATGCG 900
Qy 931 TCGAGAGCTTTCAG 990
Db 901 TCGAGAGCTTTCAG 960
Qy 991 CGTTTCATCAG 1050
Db 961 CGTTTCATCAG

also used in gene therapy. The present sequence is PAL consensus DNA.
(Updated on 07-AUG-2003 to correct OS field.)

Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;

Query Match 67.3%; Score 1629; DB 6; Length 2475;
Best Local Similarity 78.0%; Pred. No. 2.9e-249;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

37 ATGGCCCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 96
1 ATGGCCCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 60
97 CACGCGCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 154
61 CACGCGCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 120
155 GCGTGAATGGCAAGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 214
121 GCTGCTCTCTCCGACCAACCAAGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 180
215 ACCCCACGAC---GACGTCGTGAGCTCAGCGGGTACAGCTTCACCGTCTGAGCTTG 271
181 ACCCCACGAC---GACGTCGTGAGCTCAGCGGGTACAGCTTCACCGTCTGAGCTTG 240
272 TCGGCGCGCCCGCAAGGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 330
241 TCGGCGCGCCCGCAAGGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 300
331 CGCGTGAAGAGAGAGTGGCTTCTTCAAGGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 390
301 AAGATGACAAAGAGTGGCTTCTTCAAGGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 360
391 ACCACGCGCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 450
361 ACGATGCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 420
451 GCGCTCATGAGACCAAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 510
421 GCGCTCATGAGACCAAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 480
511 GAGCGCGCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 570
481 GAGCGCGCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 540
571 GTCACCTGCTCAGCGCTGAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 630
541 GTCACCTGCTCAGCGCTGAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 600
631 AACTTCTTGAACCAAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 690
601 AACTTCTTGAACCAAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 660
691 GAGGACCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 750
661 GAGGACCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 720
751 GTTCAAGCTTGAAGCAAGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 810
721 GTTCAAGCTTGAAGCAAGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 780
811 TTTGCTCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 870
781 TTTGCTCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 840
871 GCGCTCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 930
841 GCGCTCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 900
931 TCGCAGGCGCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 990
901 TCGCAGGCGCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 960

991 CCGTTCAATCCAGGCTGCTGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1050
961 CCGTTCAATCCAGGCTGCTGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1020
1051 CGCAGCTCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 1110
1021 CGCAGCTCTTCTTGGACTCGCTCGCCACCAAGCTCGCCCAAGGCTTTTACCAAGGCTCG 1080
1111 GAGCAGGAGGCAATTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1170
1081 GAGCAGGAGGCAATTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1140
1171 GAGCAGGAGGCAATTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1227
1141 GAGCAGGAGGCAATTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1200
1228 ACGACGACGCAACCGCTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1287
1201 TCGACGACGCAACCGCTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1260
1288 TTTGCTCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1347
1261 TTTGCTCTGAGGCAAGTGGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1320
1348 GCGCAGCTCAATTTCAAGGCTGAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1407
1321 GCGCAGCTCAATTTCAAGGCTGAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1380
1408 CTTTGTGCTCTGCTGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1467
1381 CTTTGTGCTCTGCTGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1440
1468 CACATGCTGCTTGAAGCTGAGCTCGCGCGCGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1527
1441 GCGCGGCGCGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1500
1528 CAGCGCGGCAAGTGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1587
1501 CAGCGCGGCAAGTGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1560
1588 ACTGCGGAGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1647
1561 ACGCGCGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1620
1648 CAGCGCGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1707
1621 CAGCGCGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAG 1680
1708 CCGACTTCTTGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1755
1681 NCGGCTCTTGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1740
1756 CTTGCTCTGAGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1815
1741 CTTGCTCTGAGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1800
1816 CTTGCTCTGAGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1875
1801 CTTGCTCTGAGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1860
1876 TCGTCTGCTGCGCA--ACGTCACCTTGAAGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGG 1933
1861 NCGTCTGCTGCGCA--ACGTCACCTTGAAGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGG 1920
1934 CCGCGGAGAGGCAATTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1993
1921 CCGCGGAGAGGCAATTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 1980
1994 CTTTGTGCTCTGCTGCGCA--ACGTCACCTTGAAGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGG 2053
1981 CTTTGTGCTCTGCTGCGCA--ACGTCACCTTGAAGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGG 2040
2054 TCGGAGGAGGCTTCTTGGCAAGTGGCTTCTTGGCAAGTGGCAAGTGGCAAGTGGCAAGTGG 2113

```
Db      2041  TCCGCGAGAGCTCGGCTCAAGAGCCCGCGGCGACGCTTCTTCTCGCAAGCAGAGG 2100
Qy      2114  AGACGATCCGAGCAAGCTCTGCGGATCTAGAGGCCATCAAGACGCGCGATCAAC 2173
Db      2101  TGACGATCGGACCAACGCTTCCGATCTACGAGGCATCAAGNCGGCGATCAAC 2160
Qy      2174  AGCTCTCTGCAAGATGCTCGGTA-AGGCCGAGCAAGCTCGCTAGAGCGCGCTC 2212
Db      2161  AGCTCTCTGCAAGATGCTCGGTAAGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 2220
Qy      2233  ACCCGAGACGAGCTTTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2292
Db      2221  NNNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 2280
Qy      2293  TGTGCTTACTCTCTGCGGCTCATCAAGCTCTGATCCGCGCTCTC 2352
Db      2281  TNNNNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 2340
Qy      2353  GGTCTCTCA-CTACAGCTGTATAGAGCTGGAATGAGTTCAGTTCGAGTTCAAAA 2411
Db      2341  NGTNNNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGN 2400
Qy      2412  AAAAAA 2419
Db      2401  NNNANAAA 2408

RESULT 8
AAD3097
ID  AAD3097 standard; DNA; 2151 BP.
XX
AC  AAD3097;
XX
DT  01-JUL-2002 (first entry)
XX
DE  R. glutinis phenylalanine ammonia lyase (PAL) mutant DNA, RM120-2.
XX
KW  Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
KW  phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW  cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KW  LCP; telecommunication; medical device; aerospace application; enzyme;
KW  biocatalyst; mutant; variant; gene; ds.
XX
OS  Rhodotorula glutinis.
XX
FH  Synthetic.
XX
FH  Key Location/Qualifiers
FT  CDS 1..2151
FT      /tag= a
FT      /product= "Rhodotorula glutinis PAL mutant"
FT      /transl_except= (pos:592..594, aa:asp)
FT      /BC_number= "4.3.1.5"
FT      replace(526, T)
FT      /tag= b
FT      replace(592..593, GG)
FT      /tag= c
FT      replace(645, G)
FT      /tag= d
FT      replace(792, A)
FT      /tag= e
FT      replace(859, T)
FT      /tag= f
FT      /tag= g
FT      replace(1619, T)
FT      /*tag= g
XX
XX  MO200210407-A1.
XX
XX  07-FEB-2002.
XX
XX  23-JAN-2001; 2001MO-US002099.
XX
XX  27-JUL-2000; 2000US-00627216.
PR
```

```
XX
XX  (DUP) DU PONT DE NEMOURS & CO E. I.
PA
XX  Gatenby AA, Sariastani FS, Tang X, Qi WM, Vannelli T;
XX
XX  WPI; 2002-280635/32.
DR
XX  P-PSDE; AAE20665.
XX
XX  Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT  polypeptide, or mutant TAL polypeptide, useful for producing
PT  para-hydroxycinnamic acid in recombinant host cell lacking cinnamate
PT  hydroxylase.
XX
XX  Example 22; Page; 139p; English.
XX
XX  The invention relates to methods for biological production of para-
CC  hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC  biocatalyst for conversion of glucose to PHCA by incorporation of the
CC  wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC  Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC  wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC  to developing a new biocatalyst for conversion of glucose to PHCA by
CC  incorporation of the wildtype PAL which possesses enhanced tyrosine
CC  ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC  plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC  Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC  recombinant techniques. The recombinantly produced PHCA may be used as a
CC  monomer for production of liquid crystal polymers (LCP). LCP may be used
CC  in electronic connectors and telecommunication and aerospace
CC  applications. LCP resistance to sterilizing radiation has also enabled
CC  these materials to be used in medical devices as well as chemical, and
CC  food packaging applications. The present sequence is Rhodotorula glutinis
CC  PAL mutant DNA. Note: This sequence is not shown in the specification,
CC  however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC  (AAD3075) in the sequence listing
XX
XX  Sequence 2151 BP; 383 A; 832 C; 563 G; 373 T; 0 U; 0 Other;
SQ
XX
XX  Query Match 49.5%; Score 1198.2; DB 6; Length 2151;
XX  Best Local Similarity 74.0%; Pred. No. 4.8e-181;
XX  Matches 1566; Conservative 1; Mismatches 524; Indels 24; Gaps 3;
XX
Qy  99  CGCCGCTCCAGCAAGTGGCTGCGGCCCACTTGGCTCTCGCGGACGCCGCGCT 158
Db  45  CGTGCATCCGCAAGAGGCTGTCAATGCGGCTTGACCACTTGAGTGCAGGCTC 104
Qy  159  CGATGACCAGCGCGGACAGTCGAGCTGATCGTGCAGAGCTCTCAGCAGCC 218
Db  105  GCACCTGCCCAACAACAGGTACAGAGTCAATCTGCAAGAGATCTGCGCGCC 164
Qy  219  CACCGAGAGCTGTGAGTTCAGCGGATACAGCTTACCTGCTGAGCTTTGCGCGC 278
Db  165  GACCGACTTCAGCTGAACTCAGCGCTACTGCTCAACTCGAGAGAGTGTCTCGCC 224
Qy  279  CGCCGCAAGGAGGAGGCTCGGCTCGAGAACGACGAGTCCGCGCAGCGTGA 338
Db  225  CGCGAGAGAGGAGGAGGCTGTCCGCTCAAGAGACGAGATCTGCTCAAAATTGA 284
Qy  339  CAAGAGCTGACTTCTTAAGGCCAGCTTCAAGAACTGAGTCTAAGAGTCAACAGCGG 398
Db  285  CAATCGTGTGAGTCTTGTGCTGCACTTCATGAGAGCTGACGGGTGACAGCTGG 344
Qy  399  TTTCGATGCTCGCGGACGACGAGATGAGATGAGTCAACCTCCAGAGGCTGAT 458
Db  345  ATTTCGATGCTCGCGGACGACGAGATGAGATGAGTCAACCTCCAGAGGCTTCT 404
Qy  459  CGAGCAGAGCTTTCGCGGATGACGCGGACGCTGCTGCTTCAAGCTGAGCGCG 518
Db  405  CGAGCAGAGCTTTCGCGGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 464
Qy  519  CTTGAGAACAGCTTTCGCGGATGCTTTCGCGGATGCTTTCGCGGATGCTTTCG 578
Db  465  TCTGAGAACTGCTTTCGCGGATGCTTTCGCGGATGCTTTCGCGGATGCTTTCG 524
```

2Y 579 GCTCAGCGTGGACACTCGGCGCGGCGCTGTCGTGCTTGAAGCGCTCAACCACTTCT 638
 2b CTTAAGCCCGGCGGCACTGGGCTGTGCGCTGTGCTGTGAGGCGCTTCAACCACTTCT 584
 2Y 639 GAACCAACCGCATCAAGCCCATCGTGTCCCTCGCGGCTTCATCTCGGCGTGGGCGACCT 698
 2b 585 CAACCAACCGCATCAACCCCATCGTGTCCCTCGCGGCGACCATCTCTGCGTGGGCGACCT 644
 2Y 699 CAGCCCGCTCTGTAACATCGCGCGCGCATCAACCGCTCAACCGCGAGCTCAAGTTACGT 758
 2b 645 CTCTCTCTCTCCACATTGACGGGCGCATGAGGCTCACCGCGAGCAAGAGTTGACAT 704
 2Y 759 TTGACACAGGGAACCGAAGATCATGTTTGCGCGCGAGGCGCATCTCGCTCTTGGTCT 818
 2b 705 GCTCAACAGGCGGCAAGAGAGATCTGTAGCGCCCGGAGGCGCATGGCGCTTCAACCT 764
 2Y 819 CGAGGCACTGTCTCTCGGCGCGAAGAGGCTGTGCTGTGTCAACGGAAGCGGCTCTTC 878
 2b 765 CGAGCCCGCTCTCGGCGCGAAGAGGCTGTGCTGTGTCAACGGAAGCGGCTCTTC 824
 2Y 879 CGCCTCGATGGGAGACCTTCAGTCTGCAAGCTGCAATGCTCTCGCTCTCGAGGCG 938
 2b 825 AGCATGATGGGCAACCTCTGCTCTGCAAGAGCAACATGCTCTCGCTCTCTCGAGTTC 884
 2Y 939 CTTGACGGCTCTCAAGGTGAGGCGCATGCTGAGCGAGGCTCTGTTGCGCGCTTCAAT 998
 2b 885 GCTACGGGCGATGAGGCTGAGGCGCATGCTGAGCGAGGCTCTGTTGCGCGCTTCAAT 944
 2Y 999 CCAAGACGTCTGCGCGCGCGACCCCGGCGAGGTGAGAGTGGCGGCAACATCGGCGAGCT 1058
 2b 945 TCACAGCTCTGCGCGCGCTCAACCGGCGAGATGAGAGTGGCGGAAACATCGGCGAGCT 1004
 2Y 1059 CTTTCCGCGCTGCTGCTTGGCGTTGAGCAAGAGAGAGGTCAAGGTCAAGAGCAAGCA 1118
 2b 1005 CTTCAAGAGGAAAGCGGCTTGTGCTGTCAACAGAGAGAGGTCAAGGTCAAGAGCAAGCA 1064
 2Y 1119 GGGCATTCTTGGCGAGAACCGCTACCGGCTCTCGCACTGCTCTGATTTCTCGGCGGCT 1178
 2b 1065 GGGCATTCTTGGCGAGAACCGCTACCGGCTCTCGCACTGCTCTGATTTCTCGGCGGCT 1124
 2Y 1179 CGTGAAGACATGATGCAAGGCGCTACTGACCTCTCGCTCGAGAAACA---CGACGAC 1235
 2b 1125 CGTCAAGCGACTCATTTCAAGGCGCGCTCTTCAACATCGAGGCGCGGCAAGTGAAGAC 1184
 2Y 1236 CGACAACCGGCTCTGCAAGTGAAGAACAGCAAGACCGGCGAGCGGCGCAACTTCAAGGC 1295
 2b 1185 CGACAACCGCTCTCATCGAGTGAAGAACAGCAACTTCAAGCGGCGGCAACTTCCAGGC 1244
 2Y 1296 GTGCGCTGTCTGATTTGATGAGAGAAACAGAGGTGCGCACTCGGCGCTCATCGGCAAGCT 1355
 2b 1245 TCCCGCTGTGGCAACACCATGGAAGAACTCGCTCGGCGCTCGCCAGATCGGCAAGCT 1304
 2Y 1356 CAACCTTCAAGAGTGAAGAGTGTCTCAAGCTGCAAGAGCGGCGGCTGCTTCTGTG 1415
 2b 1305 CAACCTTCAAGAGTGAAGAGTGTCTCAAGCGGCGATGAGAACCGGCGGCTGCTTCTGTG 1364
 2Y 1416 CTTGCTGCGAGAACCGGCTGCTCAACTATCAAGGCAAGGCTTGAACATTTCAATCGC 1475
 2b 1365 CTTGCTGCGAGAACCGGCTGCTCAACTATCAAGGCAAGGCTTGAACATTTCAATCGC 1424
 2Y 1476 TGGTTACGTTTGGAGCTGGGCAACCTTGGCAAGCGGCTCACTACTTCTGCGAGCGCGC 1535
 2b 1425 GGGGTACCTTGGAGTGGGACCTTGGCAAGCGGCTTGAAGAGATGTCACCGCGGC 1484
 2Y 1536 AGAGATGGGCAACAGAGCGGTCAACTCGTGTGCTCTCACTTCTCGGCGCGGCGACTGCGGA 1595
 2b 1485 TGAATGGGGAACAGAGGAGTCACTCGTGTGCTCTCACTTCTCGGCGCGGCGACTGCGGA 1544
 2Y 1596 GGGCAAGAGAGTCTTCTCTCTCTTCTGCGCTGCAACTGTGATGCAAGCTTCCAGGCGGT 1655
 2b 1545 GTCCAAAGAGAGTCTTCTCTCTCTCTCTGCGCAACCACTTCACTGCGGTTCGCAAGCAT 1604

QY 1656 CGACCTTCGGCGAGTAGAGCTGCATTTCAAGAAAGCATTTGACCCGCTTCTCCGACTCT 1715
 Db 1605 CGACTTCGCGGACAGAGATTGAGTTCAAGAGCAGTTGCGGCCAGGCACTGTCTGCT 1664
 QY 1716 CTTCAACAGCACTCTGGGACTGGGCTCGACGCA-----ACGCACTTGGCT 1763
 Db 1665 CATGACCAAGCACTTGGCTCTGCGCCATGACCGGCTGGAACCTTCGCGAGAGAGTCTGCA 1724
 QY 1764 CGAGTCAAGAAAGCGCTCAACAAAGCTCTGAGAGACAGACAGACGATACGCTGAGCTC 1823
 Db 1725 GAAAGTGAACAAAGCGCTGCGCAAGCGCTCGAGCAGACCACTCGTACGACTCTGCTCC 1784
 QY 1824 GCGTTGGACAGACGCTTCTCTGTAACGACACCGGCAACGTCGTGAGCTCTCTGCTTC 1883
 Db 1785 GCGTGGACAGACGCTTCTCTGTAACGACACCGGCAACGTCGTGAGCTCTCTGCTTC 1844
 QY 1884 GCGCTTGCACAGCTCAACCTTACTGCGGCTCAAGCGGTGAGAGTTCCTCGGCGAGAA 1943
 Db 1845 GTC-----GCTCTGCTCGCGCGCGCTCAAGCGCTGGAAGTTCGCGCGCGAGTTC 1895
 QY 1944 GGCATCTGCTCAAGCGGCGAGAGTTCGCAACCGCTTCTGCGAAGACCGCTCTTCGAGGC 2003
 Db 1896 GGCATCTGCTCAAGCGGCGAGAGTTCGCAACCGCTTCTGCGAAGACCGCTCTTCGAGGC 1955
 QY 2004 GCGGCGAGAGCATACCTCTGCGCGGCAAGCGGCTCTGCTGCTGCTGCGCGAGGA 2063
 Db 1956 GCGCGGCTCTGTAACCTTCTGCGCGGCACTCAAGTCTTACGCTTCTGCGCGAGGA 2015
 QY 2064 GCTCGGCTGAGAGCGCGCGCGGCGAGTGTGTCGCGCTGCAAGAGAGTTCGATCGG 2123
 Db 2016 GCTTGGCGTCAAGGCGCGCGGAGAGTCTTCTGCGGCAAGAGAGTTCGATCGG 2075
 QY 2124 GAGCAAGCTCTGCGGCACTTCAAGAGGCGCATCAAGAGAGCGGCGGATCAACAGCTCTCGT 2183
 Db 2076 CTCGAAGCTTCAAGAGTCTTCAAGAGGCGCATCAAGTGGGAGATCAACAGCTCTCT 2135
 QY 2184 CAAGTCTGCGGTA 2198
 Db 2136 CAAGTCTGCGTTA 2150

RESULT 9
 AAD33076
 ID AAD33076 standard; DNA; 2151 BP.
 XX
 AC AAD33076;
 DT 01-JUL-2002 (first entry)
 XX
 DE Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant DNA.
 KW Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KW cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
 KW LCP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst; gene; mutant; variant; ds.
 XX
 OS Rhodotorula glutinis.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2151
 FT /tag= a
 FT /product= "Rhodotorula glutinis mutant PAL"
 XX
 FN MO200210407-A1.
 XX
 PD 07-FEB-2002.
 XX
 PF 23-JAN-2001; 2001WO-US002099.
 XX
 PR 27-JUL-2000; 2000US-00627216.
 XX

(DUBO) DU PONT DE NEMOURS & CO E. I.

Gateby AA, Sarisaelani FS, Tang X, Qi W, Vannelli T;

WPI; 2002-280635/32.

P-PSDB; AAE20654.

Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL) polypeptide, or mutant TAL polypeptide, useful for producing parahydroxycinnamic acid in recombinant host cell lacking cinnamate hydroxylase.

Example 10; Page 95-98; 139pp; English.

The invention relates to methods for biological production of para-hydroxycinnamic acid (PHCA). The invention relates to developing of a new biocatalyst for conversion of glucose to PHCA by incorporation of the wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast *Rhodotorula glutinis* into *Escherichia coli* underlying the ability of the wildtype PAL to convert tyrosine to PHCA. The invention is also directed to developing a new biocatalyst for conversion of glucose to PHCA by incorporation of the wildtype PAL which possesses enhanced tyrosine ammonia lyase (TAL) activity from the yeast *Rhodotorula glutinis* plus the plant cytochrome P-450 and the cytochrome P-450 reductase into *E. coli*. Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by recombinant techniques. The recombinantly produced PHCA may be used as a monomer for production of liquid crystal polymers (LCP). LCP may be used in electronic connectors and telecommunication and aerospace applications. LCP resistance to sterilizing radiation has also enabled these materials to be used in medical devices as well as chemical, and food packaging applications. The present sequence is *Rhodotorula glutinis* PAL mutant DNA

Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

Query Match 49.5%; Score 1196.6; DB 6; Length 2151;

Best Local Similarity 74.0%; Pred. No. 8.6e-181;

Matches 155; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

99 CCACGCTCCGACCAAGTCGAGTCGCGGCCCACTTCGCTCCGCGGACGCGCGCT 158
Db CCGTCGATCCGACCAAGCGAGCTGTAAATGGCCCTCGACCAACTTCGAGTCCGAGCTC 104
QY 159 CGATGCGCAGCGCGCGCAGCAGTCGAGTCGAGATCGTGAGAGAGCTCTCCACGACCC 218
Db 105 GCACCTGCGCCACMACCCAGATCAGCAGATGACATGTCGAGAAATGCTCCGCGGCC 164
QY 219 CACGACGACGTCGTCGAGCTCAGCGGGTACAGCTCAAGCTCCGTCGATCTTGTGGCGC 278
Db 165 GACCACTCGAGCTCGAAGTCGAGGGTACTCGCTCAACTCGGAACGTCGTCGCGC 224
QY 279 CGCGCGCAAGGGCGCAGAGTCGCGTCAGAACGACGAGAGATCCGCGCAGCGCTCA 338
Db 225 CGCGAGAAAGGCGAGCGCTGTCCGGTCAGAGACGAGAGATCCGCTCAAGATTGA 284
QY 339 CAAAGAGCTCGACTTCCTCAAGGCCAGCTTCAGAAATCGCTTACGAGTCCACGCG 398
Db 285 CAATGAGTCGAGTTCGCTGCGCTCGCAACTCTCCATGAGGCTTACGCGCTCCGACTGG 344
QY 399 TTTGAGTGGCTCGCGCAGACGAGAGATGACAGCTCAGCTCCAGAAAGCGCTCAT 458
Db 345 ATTTGGCGGATCCGAGACACCCGACGAGAGAGCATTCGCTCCAGAGGCTTCTCT 404
QY 459 CGAGACGAGCTTCGCGCGTGAAGCGCAGTCCGCTCTCTTACGCGTCGAGCGCG 518
Db 405 CGAGACGAGCTTCGCGCGTGTTCCTTCCTGCTTCGACTCGTTCGCGCTCGCGCGCG 464
QY 519 CCTGAGAGACGCTTCGCTCGAGTGTCCGCGGCGAGAGATGATCCGCGTCAATC 578
Db 465 TCTGAGAACTCGCTTCCTTCGAGGTTTTCGCGGCGACATACATCGCGCTCAACG 524
QY 579 GCTACGCGTGGCACTTCGCGCGCTCGCTGCTCTTGAAGCGCTCAACATTTCTT 638

Db 525 CTTGACCGCGGCACTCGGCTGTCCGCTCGTCTCTCGAGGCGCTCAACATTTCT 584
QY 639 GAACACGCGATCAGGCCATGTCCTCCCTCCGCGCTCAATCTCGAGTCGAGCGACT 698
Db 585 CAACCAAGGACATCAACCCCATGTCCTCCCTCCGCGGACCACTCTCTGCTCGGCGCACT 644
QY 699 CAGCCGCTCTGTATCATCCCGGCGCATCAACGCTACCCGACCTCAAGGTTCACT 758
Db 645 CTCTCTCTCTCTCAATTTGACGCGCATCAGCGCTACCCGACGACGAGGTGACGT 704
QY 759 TTTGACGAGGAAACCGAGAGATCATGTTGCGGAGAGCATCTCGCTCTTGGTCT 818
Db 705 CCTCAACGAGGCAAGAGAAATCTGTACGCGCGGAGCGATGGCGCTTTCAACT 764
QY 819 CGAGCAATGCTCTCGGCGCGGAGAGAGGTTCTGATCAACGGAACGCGCTCTC 878
Db 765 CGAGCGCTGCTCTCGGCGCGGAGAGAGGTTCTGATCAACGGAACGCGCTCTC 824
QY 879 CGCCTGATGGCGACCCCTGAGTTCGACAGCTCGACATGCTCTCTCTCTCGAGCG 938
Db 825 ACATCGATGGCAACCTCGCTTGCACGACGACATGCTCTCTCTCTCGAGTCT 884
QY 939 CTTGACGCTCTCAACGCTGAGGCGCATGTCGCGCAGACGAGGCTCGTTCGCGCTTCA 998
Db 885 GCTACGCGCATGACGCTGAGAGATGTCGCGCAGCGGCTCTCTCAACCTTCTCT 944
QY 999 CACGACGCTTCGCGCGCGGACCCCGGCGAGTTCGAGTTCGCGGCAACATCCGACGCT 1058
Db 945 TCAGACCTCAAGCGGCGCTCAACCGAGATCGAAGTTCGCGGAAACATCCGCAAGCT 1004
QY 1059 CTTTCGCGCTGCTGCTTGGCTTGGAGCAAGAGAGAGTCAAGTCAAGAGAGAGCA 1118
Db 1005 CTTGAGAGAGAGCGGCTTCTGCTGTACCACTGAGAGAGATCAAGTCAAGAGAGAGCA 1064
QY 1119 GGGATCTTTCGCGAGAGACCGCTACCCGCTCCGACAGTCTCGCTCTCGGCGCGCT 1178
Db 1065 GGGATCTTTCGCGAGAGACCGCTACCCGCTTGGCGAGCTCTCTCTCTCTCTCTCTCT 1124
QY 1179 GGTGAGAGACATGATGAGCGCTACTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1235
Db 1125 GGTGAGAGACATGATGAGCGCTACTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1184
QY 1236 GACCAACCGGCTCTCGAGCTCGAGAACAGCGGAGCGGCGGCACTTCGAGCG 1295
Db 1185 GACCAACCGGCTCTCGAGCTCGAGAACAGCGGAGCGGCGGCACTTCGAGCG 1244
QY 1296 GTGCGCTGTCTGATTTGATGAGAGAGACAGGCTCGACCTGCTCTCTCTCTCTCTCT 1355
Db 1245 TGCAGCTGTGAGCAACATGAGAGAGACTCGCTTCGCGCTTCGCGCATGAGGAGCT 1304
QY 1356 CAATTCAGCAGTGCACGAGTGTCAACGCTGTCATGAACCGGCGGCTCTCTCTCTCT 1415
Db 1305 CAATTCAGCAGTGCACGAGTGTCAACGCTGTCATGAACCGGCGGCTCTCTCTCTCT 1364
QY 1416 CTTGCTGCGAGAGATCCGCTGCTCAATATCAAGGCAAGGCTTGAATTCACATGCG 1475
Db 1365 CTTGCTGCGAGAGATCCGCTGCTCAATATCAAGGCAAGGCTTGAATTCACATGCG 1424
QY 1476 TGTTCAGCTTGGAGCTGGAGCTGGAGCTGCAACCGGCTCACTCACTTCGTCACAGCG 1535
Db 1425 GGGTACACTTGGAGTGGAGACTGCGCAACCTGTGAGAGAGCAATGTCACAGCGCG 1484
QY 1536 AGAGATGGGCAACAGGCGGTCAATCGCTGCTCTCACTTCGCGCGCGGCACTGCGCA 1595
Db 1485 TGAATGGGCAACAGGCGGTCAATCGCTGCTCTCACTTCGCGCGCGGCACTGCGCA 1544
QY 1596 GGCACAGAGCTCTTCT 1655
Db 1545 GTCCAGAGAGCTCTTCT 1604
QY 1656 CGACTTCGCGGAGAGAGCTGCACTTCAAGAGAGTTCGACCGGCTTCTCCGAGCTCT 1715
Db 1605 CGACTTCGCGGAGAGAGCTGCACTTCAAGAGAGTTCGCGGCTTCTCTCTCTCTCTCTCT 1664

YY		1716	CCTCCAGAGACGACCTCGGCACATGGCCTCGACGTCGA-----ACCGACTTGGCGCT	1766
YJ		1655	CATGACCAGCACTTTTGACTTCGCCCATGACCAGCTCGAATCTGGCGCACAGAGCTTCGTGA	1724
YI		1764	CGAGGTCAGAAGGCGCTCAACAAGCGCTCTCGACAGACGACGATACGACTTCGAGGCC	1823
YH		1725	GAAAGGTGAACAAGACGCTCGCCAAGCGCGCTGAGACAGACCAACTGTATACGACTCTGTCGCC	1784
YD		1824	GCGCTGGACGACGCGCTTCTGTACGGGACCGGACCGGTGGTGGAGCTCCCTGTGTCCTC	1883
YC		1785	GCGCTGGACGACGCGCTTCTGTACGGGACCGGTGGTGGAGCTCCCTGTGTCCTC	1844
YB		1884	GCCCTCTGCCAACGTCACCCCTTACTGTCCGCTGACGCGGTGAAGGTTGCTTCGSCGAGAA	1943
YA		1845	GTC-----GCTCTGCTCGCCGCGGTCAACGCTGGAAGGTGCGCGCGCGAGTC	1895
YX		1944	GCGCATCTCGCTCAACGCGCGAGAGTGGGAAACCGCTTTCGACAGACGCGCTTTCGACAGGC	2003
YV		1896	GCGCATCTCGCTCAACGCGCGAGAGTGGGAAACCGCTTTCGACAGAGCTTCGTC	1955
YU		2004	GCCGCGCAGCAGTCATCTCTCGCCGCGACGCGCGCTCTGTACTGTGCTGCGCGAGGA	2063
YT		1956	GCCCCGCTCTGTACTCTCTCGCCGCGACATCATGATCTTCAACGCTTCGTCGCCGAGGA	2015
YS		2064	GCTCGCGCTGACGAGGCGCGCGCGGAGAGTGTGTGGCGGTGGACGAGAGAGATCGG	2123
YR		2016	GCTTGGCGTCAGGCGCGCGCGGAGAGCTCTTCTCGGCAAAGAGGTGACGATCGG	2075
YQ		2124	GAGCAACGTCCTCGCATCTACGAGGCCATCAMAGACGCGCATCAACACGTCCTCGT	2183
YP		2076	CTCGAAGCTCTCAAGATCTACGAGGCCATCAAGTCGGGACAGATCAACAACGTCCTCT	2135
YN		2184	CAAGATGCTCGCGTA	2198
YM		2136	CAAGATGCTCGTTA	2150
YL				
YK				
YJ				
YI				
YH				
YD				
YC				
YB				
YA				
YX				
YV				
YU				
YT				
YS				
YR				
YQ				
YP				
YN				
YM				
YL				
YK				
YJ				
YI				
YH				
YD				
YC				
YB				
YA				
YX				
YV				
YU				
YT				
YS				
YR				
YQ				
YP				
YN				
YM				
YL				
YK				
YJ				
YI				
YH				
YD				
YC				
YB				
YA				
YX				
YV				
YU				
YT				
YS				
YR				
YQ				
YP				
YN				
YM				
YL				
YK				
YJ				
YI				
YH				
YD				
YC				
YB				
YA				
YX				
YV				
YU				
YT				
YS				
YR				
YQ				
YP				
YN				
YM				
YL				
YK				
YJ				
YI				
YH				
YD				
YC				
YB				
YA				
YX				
YV				
YU				
YT				
YS				
YR				
YQ				
YP				
YN				
YM				
YL				
YK				
YJ				
YI				
YH				
YD				
YC				
YB				
YA				
YX				

```

PD      07-FEB-2002.
PE      23-JAN-2001, 2001MO-USO02099.
PR      27-JUL-2000; 2000US-00627216.
XX      (DUPO ) DU PONT DE NEMOURS & CO E. I.
PI      Gatenby AA, Sariastani FS, Tang X, Qi WW, Vannelli T,
DR      WPI; 2002-280635/32.
XX      P-PSDB; AAE20668.
PT      Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT      polypeptide, or mutant TAL polypeptide, useful for producing
PT      parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PS      hydroxylase.
PS      Disclosure; Page; 139pp; English.
XX      The invention relates to methods for biological production of para-
XX      hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC      biocatalyst for conversion of glucose to PHCA by incorporation of the
CC      wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC      Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC      wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC      to development of a new biocatalyst for conversion of glucose to tyrosine
CC      incorporation of the wildtype PAL which possesses enhanced tyrosine
CC      ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC      plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC      Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC      recombinant techniques. The recombinantly produced PHCA may be used as a
CC      monomer for production of liquid crystal polymers (LCP). LCP may be used
CC      in electronic connectors and telecommunication and aerospace
CC      applications. LCP resistance to sterilising radiation has also enabled
CC      these materials to be used in medical devices as well as chemical, and
CC      food packaging applications. The present sequence is Rhodotorula glutinis
CC      PAL mutant DNA. Note: This sequence is not shown in the specification,
CC      however it is constructed based on the PAL DNA shown as SEQ.ID.NO.7
CC      (AAD33075) in the sequence listing
SQ      Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;
SQ      Query Match          49.5%; Score 1196.6; DB 6; Length 2151;
SQ      Best Local Similarity 74.0%; Pred.No. 8.6e-181;
SQ      Matches 1655; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

Query Match          49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred.No. 8.6e-181;
Matches 1655; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

99 CGCGGCTTCGACCAAAGTGCGGTGCGGGGCCCACTTCGGCTCTCCGGCCGACGCCCGGCT 158
    |||||
45 CGTCCCATCCGCAAAAGCAGCGGTGCATATGAGGCGCTCGACCAACTTCGACGTGCGAGGCTC 104
    |||||
159 CGATGGCACCGCGCGGACCAAGTCGACAGTCTGTGAGAGAGTCTCTAGCGAACCC 218
    |||||
105 GCACCTGCCACAACCAGATCACGCAAGTCACGCAATGTCGAAGAAGATGCTCGCGCGCC 164
    |||||
219 CACCGACGACGTCGTGAGACTTAGCGGGTAAGCTCAACGTCGTGACGTTTGCGGCGC 278
    |||||
165 GACCGACTCGACGCTCGAATCGAAGCTGACTACGCTCAACTCGGAGACGTCGTTCGAGC 224
    |||||
279 GCGCCGCAAGGGGGCGCAGAGTCCGCGTCGAAGAAGACGACAGAGATCCCGACGCGTTCGA 338
    |||||
225 CGCGAGAAAGGGCAGGCGCTGTCCGCGTCAAGSACGACGAGATCCGCTCAAAATTGA 284
    |||||
339 CAAGAGCGTCGACTTCCTCAAGGCCAGCTTCAGAACTCGGTTACGAGTCCACAGGG 398
    |||||
285 CAATTCGCGTCGAGTCTTCGCGCTCGCAACTCTCCATGAGCGCTCTCGCGCTCAGCATGG 344
    |||||
399 TTTCGCTGCTGCGCGCAGACGAGAGACTGAGAGATCAGATCAAGACCTCCAGAAAGGCGTCAT 458
    |||||
345 ATTTGGCGGATCCGAGACACCCGACGCGAGAGACCCATCTCGCTCCAGAAAGGCTCTCT 404
    |||||
459 CGAGACACGAGCTCTCGGCGGTACCCGACGTCGCTCGCTCTTCAGCGTTCGACGCGG 518
    |||||

```


PD 15-FEB-2001.
XX 03-AUG-2000; 2000WO-US021156.
XX 06-AUG-1999; 99US-0147719P.
XX (DUPO) DU PONT DE NEMOURS & CO E. I.
XX Tang X, Vannelli TM, Qi W, Sarisiani S, Gatenby AA;
XX P-PSDB; AAE16390.
XX WPI: 2002-121549/16.
XX Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia
PT -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT activity.
XX
XX Claim 13; Page 67-68; 75pp; English.

CC The invention relates to a method for producing para-hydroxycinnamic acid
CC (PHCA). The method comprises converting cinnamate to PHCA by converting
CC glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC route or by generating a new biocatalyst possessing enhanced tyrosine
CC ammonia-lyase (TAL) activity. The method is useful for the biological
CC production of PHCA which is useful as a monomer for the production of
CC Liquid Crystal Polymers (LCP), where the LCP is used in electronic
CC connectors, telecommunications and aerospace applications. LCP is also
CC useful in medical devices, as well as chemical, and food packing
CC applications due to its resistance to sterilizing radiation. The present
CC sequence is a DNA encoding red yeast, Rhodotorula glutinis
CC (Rhodospiridium toruloides) mutant PAL derived by substituting Ile at
CC position 540 with Thr. Mutant PAL enzyme possesses enhanced TAL activity
XX
XX Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

Query Match 49.5%; Score 1196.6; DB 6; Length 2151;
Best Local Similarity 74.0%; Pred. No. 8.6e-181;
Matches 1665; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 GCGCGCTCGACCAAGTCGCGTGGCGGCCCACTCGGCTCGCGCGCAAGCGCGCT 158
DB 45 CTTCCGATCCGCAAGGCAAGCTGTATAGCGCTCGACCACTCGCACTGCGAGCTC 104
QY 159 CGATGCGCAAGCGCGCGCAAGTCGAGCTCGATCGTGAAGAGCTCTTCAAGCAACC 218
DB 105 GCACCTGCCCAACCAACCGATGCAAGCTCGATCGTGAAGAGTGTCTCGCGCGCC 164
QY 219 CACCGACGAGCTGTGAGCTTCAAGGCTACAGCTTACCGTCCGTACGTTGCGCGCG 278
DB 165 GACGACCTGACGCTTGAAGCTGAGGCTACTCGCTCACTCGGAAGCTGTCTCGGC 224
QY 279 CCGCGCAAGGCGCGCGAGGCTCGCGTCCAGAACGACGAGAGATCCGCGACGCTCGA 338
DB 225 CCGAGAGAGGCGCGAGGCTGTGCGTCAAGGACAGGAGATCCGCTCAAGATGA 284
QY 339 CAAAGGCTGACTCTTCTCAAGGCGCGAGCTTCAAGCTGTGCTTCAAGAGTCAACGCG 398
DB 285 CAAATGCTGAGCTTCTTCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 344
QY 399 TTTTGGTGGCTGGCGCGACGAGAGCTGAGATGACAGCTTCAAGAGGCGCTCAT 458
DB 345 ATTGGCGATCGAGACACCGGACCGAGAGAGCTTCTGCTTCAAGAGGCTTCTCT 404
QY 459 CGAGCAACGACTTGGCGCGTGAAGCGCAAGCTGCGCTCTTCAAGAGTGGAGCGCG 518
DB 405 CGAGCAACGACTTGGCGCGTGAAGCGCAAGCTTCTGCTTCAAGAGGCTTCTCT 464
QY 519 CCTGAGAGCAAGGCTTCTGCTTGAAGTGTGCGCGCGCGAGTGTGCTTCAAGAGTCTC 578
DB 465 TCTGAGAGCAAGGCTTCTGCTTGAAGTGTGCGCGCGCGAGTGTGCTTCAAGAGTCTC 524
QY 579 GCTCAGCGCTGAGCACTGAGCGCTGCGCTGCTGCTTGAAGGCGCTTCAAGAGTCTT 638

DB 525 CTTGACCCCGGCGCACTCGGCTGTCCGCTCTGCTGTCTGAGAGGCTCAACCACTTCT 584
QY 639 GAACCAACCGATCAAGCGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 698
DB 585 CAACACGAGATCAACCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 644
QY 699 CAGCGCGCTTCTGAGCACTGCGCGCGCATCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCT 758
DB 645 CTTCTCTCTCTCTCACTTGAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCT 704
QY 759 TTTGACAGAGGAGAACCGAAGATCATGTTTGGCGCGAGGCGCATCTCGCTTGGTCT 818
DB 705 CGTCCACAGAGGAGAACCGAAGATCATGTTTGGCGCGAGGCGCATCTCGCTTGGTCT 764
QY 819 CGAGGAGCTGCTTCTGCGCGCGCGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878
DB 765 CGAGCGCTGCTTCTGCGCGCGCGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
QY 879 CGCTGAGTGGCGCACTCTGAGTGTGACAGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 938
DB 825 AGCATGATGGCGCACTCTGCTGCTGCAAGCAAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
QY 939 CTTGAGGCTTCTGAGTGTGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 998
DB 885 GCTCAGCGCTTCAAGGCTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
QY 999 CCAGAGCTTGGCGCGCGCGAGCGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1058
DB 945 TCAGAGCTTCAAGGCTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
QY 1059 CTTTCCGCTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
DB 1005 CTTGAGAGGAGAGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
QY 1119 GGGCATTTTGGCGAGAGCGCTTACCGCTCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1178
DB 1065 GGGCATTTTGGCGAGAGCGCTTACCGCTCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1124
QY 1179 CTTGAGAGGAGTGTGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
DB 1125 CTTGAGAGGAGTGTGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
QY 1236 CGACACCGCTTCTTCAAGCTTCAAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295
DB 1185 CGACACCGCTTCTTCAAGCTTCAAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1244
QY 1296 GTGCGCTGTGCTGAGTGTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
DB 1245 TGGCGCTGTGCGCAACGATGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
QY 1356 CAATTCAGAGTGTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
DB 1305 CAATTCAGAGGAGTGTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
QY 1416 CTTGCGGCTGAGAGAGCGCTGCTCACTATCAAGGAGGCTTGAAGTGTGAGTGTGAGTGTGAGTGTGAG 1475
DB 1365 CTTGCGGCTGAGAGAGCGCTGCTCACTATCAAGGAGGCTTGAAGTGTGAGTGTGAGTGTGAGTGTGAG 1424
QY 1476 TGTTCAGCTTGTGAGTGTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1535
DB 1425 GGTGTACACTGTGAGGTTGGAGACTGTGCAACCTGTGAGCGAGTGTGAGCGAGTGTGAGCGAG 1484
QY 1536 AGAGTGTGAGCAACGAGCGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1595
DB 1485 TGAAGTGTGAGCAACGAGCGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1544
QY 1596 GAGCAACAGCTTCTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
DB 1545 GTTCAAGAGAGTCTTCTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
QY 1656 GCACTTCCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1715

```

Db      1605 CGACTTGGCGGACCGAGTTGAGTTTCAAGAGAGTTCGGCCCGACCATCTGTCTGCT 1664
Qy      1716 CCTCCAGCAGCACTTGGACATGGCTTCACTCA-----ACGCACTTGGCT 1763
Db      1665 CATGACCGACCTTTGGCTCCGCCATGACCGGCTCGAAGCTCGCGAGCTCTGCGA 1724
Qy      1764 CGAGTCAAGAGGCGCTCAACAGAGTCTGAGGAGACGAGACGTTCAGCTCGAGCC 1823
Db      1725 GAAGGTGAACAGACGCTGCCAAGCGCTCGAGAGACCAACTGTGACCTCTGCTCC 1784
Qy      1824 GCGCTGGACAGACGCTTCTGACCGGACCGGACCGCTGTCGAGCTCTCTGCTCTC 1883
Db      1785 GCGCTGGACAGACGCTTCTCTGCGCGGACCGCTCTGAGGCTCTCTGCTGAC 1844
Qy      1884 GCGCTTGGCAAGTCAAGCTTACGCGTCAAGCGGTGAAGGTTGCTTGGCGGAGAA 1943
Db      1845 GTC-----GCTTCGCTGCGCGGTCAAGCTGGAAGTTCGCGCGCGAGTC 1895
Qy      1944 GGGCATCTGCTCAACGCGGAGTGGGAAACCGCTTTCGAGAGACGCGCTTTCGACAGC 2003
Db      1896 GGGCATCTGCTCAACCGCGGAGTGGGAAACCGCTTTCGAGAGACGCGCTTTCGACAGC 1955
Qy      2004 GCGGCGGACGCACTACTCTTCCCGCGACGCGCTCTGACTGCTTCTGCGCGAGGA 2063
Db      1956 GCGGCGGCTCTGTAAGCTTCTGCGCGGACCTGATCTCTACGCTTCTGCGCGAGGA 2015
Qy      2064 GCTCGGCTGTCAGCGCGCGCGGAGTGTGCTGCGCGTGGAGAGAGATCGG 2123
Db      2016 GCTTGGCGTCAAGCGCGCGCGGAGAGTCTCTCTCGCAAGAGAGTGAATCGG 2075
Qy      2124 GAGCAAGTCTCGCGCTTACGAGGCGATCAAGAGCGCGCATCAACAGCTCTCGT 2183
Db      2076 CTGGAAGCTCTCAAGATCTACGAGGCGATCAAGTGGGCGAGATCAACAGCTCTCT 2135
Qy      2184 CAAGATGCTCGCGTA 2198
Db      2136 CAAGATGCTCGCTTA 2150

RESULT 12
ABSS6543
ID ABSS6543 standard; DNA; 2151 BP.
XX
AC ABSS6543;
XX
DT 27-JAN-2003 (first entry)
XX
DE R. glutinis DNA encoding PAL/TAL naturally occurring variant.
XX
KW ds; gene; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase;
KM PAL; para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
KM tyrosine; cinnamic acid; PHCA; liquid crystal polymer.
XX
OS Rhodotorula glutinis.
XX
FH Key Location/Qualifiers
FT CDS 1..2151
FT /tag= a
FT /product= "PAL/TAL"
XX
XX WO200290523-A2.
XX
XX 14-NOV-2002.
XX
XX 03-MAY-2002; 2002WO-US018551.
XX
XX 04-MAY-2001; 2001US-0288701P.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Qi MW, Sariaelani FS, Tang X;
XX
XX WPI, 2003-058928/05.

```

```

DR      P-PSNB; ABG71255.
XX
XX Novel recombinant host for producing para-hydroxycinnamic acid and
PT tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT gene encoding phenylalanine hydroxylase activity.
XX
XX Example 4; Page 66-67; 69pp; English.
XX
XX The invention relates to a recombinant host comprising at least one gene
CC encoding a tyrosine ammonium lyase (PAL) activity, and at least one gene
CC encoding a phenylalanine hydroxylase (PH) activity. Also included are:
CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC providing the recombinant host; and (2) growing the recombinant organism
CC in the presence of a fermentable carbon substrate and thus producing para-
CC hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC providing a recombinant organism comprising at least one gene encoding a
CC phenylalanine hydroxylase activity; and (2) growing the recombinant
CC organism in the presence of a fermentable carbon substrate to produce
CC tyrosine. The recombinant host cell is used for producing para-
CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC monomer for the production of liquid crystal polymer. The present
CC sequence encodes a naturally occurring variant of the R. glutinis
CC phenylalanine ammonia-lyase, PAL, which has TAL activity (PAL/TAL).
XX
SQ      Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;
Query Match 49.5%; Score 1196.6; DB 7; Length 2151;
Best Local Similarity 74.0%; Pred. No. 8.6e-181;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
Qy      99 GCGCGCTCCGACCAAGTGGCTGGGCGCCCACTTGGGCTTTCGCGGACGCGGCGCT 158
Db      45 GGTGCGATCCGCAAGAGGCTGATGCGCTCGACCAACTCGCAGTGCAGAGCTC 104
Qy      159 CGATGGCGAGCGCGCGGACAGTTCGACGCTCGAGATTCGTCAGAGACTCTCAGGAGCC 218
Db      105 GCACCTCCGCAACCAAGGTCAAGGTCAGAGTGCAGATTCGAGAAATCTGCGCGCGC 164
Qy      219 CACCGACGAGCTGCTGAGCTCAGCGGATACAGCTTACCGTCCGTGAGTGTGCGGCG 278
Db      165 GACCGACTCGACGCTGAACTCGACGCTACTGCTCAACCTCGAGAGATGCTGTCTCGCG 224
Qy      279 GCGCCGCAAGGGCGGAGGCTCGGCTCGCAAGACGACGAGATCGGCGACGCGGTCA 338
Db      225 GCGAGAGAGGGGAGGCTGCTCGCGCTCAAGGACACCGAGATCGGCTCAAGATTGA 284
Qy      339 CAAGAGGTGAGCTTCTCAAGGCCAGCTTCAGACTCGGTCTACGGAGTCAACACCGG 398
Db      285 CAATTCGATGAGTTCTTGGCTCGCACTTCTCAAGAGCTTACGAGGCTCAACACTGG 344
Qy      399 TTTGGTGGCTCGGCGGACACGAGACTGAGATGCACTCAAGCTTCAGAGAGCGCTCAT 458
Db      345 ATTGGCGGATCCGCAAGACCCGCAACCGAGAGCGCATTCGCTCAAGAGCTCTCT 404
Qy      459 CGAACACCACTCTGCGCGGCTGACGCGGACGCTCGCTGCTTCAAGGTGAGAGCGG 518
Db      405 CGAGCACCACTCTGCGGCTGCTCTCTTGTGCTGCACTGTTCTGCTGCGCGCGG 464
Qy      519 CTTGAGAACACGCTTCTGCTGAGTCTGTCGCGGCGGACGATGATCCGCTCAAGTTC 578
Db      465 TCTGAGAACTGCTTCTCTGAGGTTGTTGGCGCGGCGCAAGATCCGCTCAAG 524
Qy      579 GCTGAGCGGTGGCACTCGGCGCTCGGCTCTGCTTGAAGGCTCAACAACTTCTT 638
Db      525 CTTGACCCCGGCGACCTGCTGCTGCTGCTGCTGCTGCTGAGGCGCTCAACAACTTCT 584
Qy      639 GAACCAACGATCAGCGGACGATCGTCCGCTCGCGGCTCATCTCGGCTCGGCGGAGCT 698
Db      585 CAACCAACGATCAGCGGACGATCGTCCGCTCGCGGCGCACTCTGCTGCTGCGGAGCT 644
Qy      699 CAGCGCGCTTCTGATCATGCGCGGCGCATCACCGGTCAACCGGACGTGAAGTTCACT 758

```


Db CTTCTCTCTCTCTCTCAATTGACAGGCGCATCAGCGCTCACCCGGAAGCAAGTGCAGCT 704
2Y TTTGGACAGAGGAACCGAAGATCATTTTGGCCGAGAGCCATCTCGCTCTTTGCTCT 818
Db CGTCACAGAGGACAGAGAGATCTGTACGCCGGAGGCAATGGCGCTTTCAACCT 764
2Y CGAGGCACTCTCTCTCGGCGCGAAGAGAGGCTGTGTCTGTCAACGGAACGCGCTCTC 878
Db CGAGCCCTCTCTCTCTCGGCGCGAAGAGAGGCTGTGTCTGTCAACGGAACGCGCTCTC 824
2Y CGCTCGATGAGCGACCTCACTGTGCAAGACTCGACATGCTCTCGCTCTCTCTGAGGC 938
Db AGCATCGATGAGCGACCTCTGTGCAAGACTCGACATGCTCTCTCTCTCTGAGGC 884
2Y CTGACGCGCTCTCAAGGTGAGGAGCATGCTGCGACAGAGGCTCTCTCTCTCTCTCTCT 998
Db GCTCAGGCGCTCAAGGTGAGGAGCATGCTGCGACAGAGGCTCTCTCTCTCTCTCTCT 944
2Y CCACGACGCTCTGCGCGCGCGACCCCGGCAAGGTGAGGCTGCGCGCAACATCCGACGCT 1058
Db TCAGACGCTCAAGGCGCGCTCAACCGACAGATCGAAGTGGCGGAAACATCCGCAAGCT 1004
2Y CTTTCCGCGCTCTGT 1058
Db CTTGAGGAGAAACCGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1064
2Y GGGCATCT 1178
Db GGGCATCT 1124
2Y GGTGAGGACATGATGACGCT 1235
Db GGTGAGGACATGATGACGCT 1184
2Y CGACACCGCGCT 1295
Db CGACACCGCGCT 1244
2Y GTGCGCT 1355
Db GTGCGCT 1304
2Y CACTTTCAAGAGTGCACCGAGTGTCTCAAGCTGTCTCAAGCTGTCTCAAGCTGTCTCA 1415
Db CACTTTCAAGAGTGCACCGAGTGTCTCAAGCTGTCTCAAGCTGTCTCAAGCTGTCTCA 1364
2Y CTTGCGCGCGAGAGCGCGCT 1475
Db CTTGCGCGCGAGAGCGCGCT 1424
2Y TGGCTTACGCTTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1535
Db TGGCTTACGCTTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1484
2Y GAGGATGAGGACACGAGCGGTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1595
Db TGGATGAGGACACGAGCGGTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1544
2Y GAGGATGAGGACACGAGCGGTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1655
Db GAGGATGAGGACACGAGCGGTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1604
2Y GAGGATGAGGACACGAGCGGTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1715
Db GAGGATGAGGACACGAGCGGTCACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1664
2Y CTTGCGCGCGAGAGCGCGCT 1763
Db CTTGCGCGCGAGAGCGCGCT 1724
2Y CTTGCGCGCGAGAGCGCGCT 1823
Db CTTGCGCGCGAGAGCGCGCT 1784

QY 1824 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 1883
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 1844
QY 1884 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCT 1943
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 1895
QY 1944 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCT 2003
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 1955
QY 2004 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCT 2063
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 2015
QY 2064 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCT 2123
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 2075
QY 2124 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCT 2183
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 2135
QY 2184 GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCT 2198
Db GCGGTGACAGAGCGCTTCTCTGTAAGGACCGGACCGGTGTGAGCTCTCTCTCTCTCT 2150

RESULT 13
ADE36055
ID ADE36055 standard; DNA; 2151 BP.
XX
XX ADE36055;
XX
XX 29-JAN-2004 (first entry)
XX
XX Rhodospiridium glutinis EPR8km-6 mutant PAL encoding DNA SEQ ID NO:3.
XX
XX recombination; nucleic acid template; recombined polypeptide;
XX
XX altered property; combined gene; gene; ds.
XX
XX Rhodospiridium.
XX
XX WO2003072743-A2.
XX
XX 04-SEP-2003.
XX
XX 26-FEB-2003; 2003WO-US005708.
XX
XX 26-FEB-2002; 2002US-0360279P.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX
XX Milano J, Tang X;
XX
XX PI
XX
XX WPI: 2003-721766/68.
XX
XX 2-PSDB; ADE36056.
XX
XX Recombination of nucleic acid templates; useful e.g. for producing
XX PT enzymes with altered properties, is based on cyclic extension of unpaired
XX PT primers.
XX
XX Example 2; SEQ ID NO 3; 481bp; English.
XX
XX The present invention describes a method for the recombination of nucleic
XX CC acid (NA) templates. The method comprises: (a) providing at least two
XX CC double-stranded templates (T1, T2) having different 5' and 3' regions in
XX CC their sense strands; (b) contacting the templates with at least one each
XX CC of forward and reverse primers that anneal, respectively, only to the 3'-
XX CC region of the antisense strand of T1 and to the 3'-region of the sense
XX CC strand of T2; (c) extending the primers by no more than 1000 nucleotides

CC (nt): (d) separating extended primers from their templates; (e)
CC reannealing the extended primers, to either template; and (f) repeating
CC cycles of steps (b)-(e) until at least one full-length extension product,
CC i.e. a recombinant of T1 and T2, is obtained. Also described: (1) a
CC method starting from at least one each of antisense and sense single-
CC stranded templates; (2) generating a recombinant polypeptide (I) with
CC altered properties by expressing a recombinant polypeptide (I) with
CC processes and screening the polypeptides formed for properties different
CC from those of polypeptides from either of the templates; and (3) (I)
CC produced by method (2). The methods are useful for the recombination of
CC nucleic acid templates of interest, and generation of a recombinant
CC polypeptide having altered properties. The method can be used to make
CC combined genes that express proteins e.g. enzymes, cytokines, growth
CC factors, viral proteins or microbial antigens with altered properties.
CC e.g. stability activity or specificity. The present sequence is used in
CC the exemplification of the present invention.

SQ Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

Query Match 49.5%; Score 1196.6; DB 9; Length 2151;

Best Local Similarity 74.0%; Pred. No. 8.6e-181; Mismatches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CGCGCTCGACCACTGCTGCGAGGCCCACTTGGCTCTCGCGCGACGCCGCGCT 158
DB 45 CCGCGATCCGCAAGACAGGCTGTCAATGGCGCCCTCGACCAACTCGAGTCCGAGCTC 104
QY 159 CGATGGCCACGCCCGCGACCAAGTCCGACTCGAATCTGTGACAGAGCTCTCAAGCACCC 218
DB 105 GCACTGCCCCAACACCCAGGTCAACGAGGTGCACTCGTCAAGAAATGCTCGCGCGCC 164
QY 219 CACCGACGAGCTGCTGAGCTGAGCGGTTAGAGCTCAAGCTCCGCTGAGCTGCGCGC 278
DB 165 GACCGACTCGAGAGCTGAACTGACGGTACTCGGTCAACTCGGAAACGTCGCTGCGC 224
QY 279 CGCGCCGAAAGGCGCGAGAGTCCGCTCGCAAGACGACGAGATCCGCGACGCTCGA 338
DB 225 CGGAGGAAAGGCGAGAGCTGTGCTCGCTCAAGACGAGACGAGATCCGCTCAAGATTGA 284
QY 339 CAAGAGCTGCACTTCTCAAGAGCCAGCTTCAGAACTCGGCTCAGGAGTCAACAGGG 398
DB 285 CAATCGTGCAGTTCTTGGCTGCGACTCTCCATGAGGCTCTACGGCTCAAGACTG 344
QY 399 TTTGGTGTGCGCGCGACAGAGACTGAGATGCAATGCAAGCTCCAGAAAGCGCTCAT 458
DB 345 ATTGGCGGATCCGAGACACCCGACCGAGAGACGCAATCGCTCAGAAAGCTCTCT 404
QY 459 CGAGACCAAGCTGTGCGGCGTGAAGCGCGACGTCGCTCGCTTCCGCTCGGACGCGG 518
DB 405 CGAGACCAAGCTGTGCGGCTGTTCTTCTGCTTGAATCGTCTCGCTCGGACGCGG 464
QY 519 CCTGAGAACAGCTTTCGCTGAGGTGTCGCGGCGCATGATGTCATCCGCTCAACTC 578
DB 465 TCTCGAAGACTGCTTCCCTCGAGTTGTTCCGGGCGCATGACATCCGCTCAACAG 524
QY 579 GGTCAAGGTGGCCACTGCGCGCTGCGCTGTGTCTTGAAGCGCTCAACTTCTT 638
DB 525 CTTGACCGCGCGCACTGCGCTGTGCTGTGCTGTGCTGAGCGCTCAACAACTTCT 584
QY 639 GAAACACGCAATCAGCCCATGCTCCCTCGCGGCTCCATCTCGGCTCGGAGCACT 698
DB 585 CAACACGCGCATACCCCATGCTCCCTCGCGGACCATCTCTGCGTCCGAGCACT 644
QY 699 CAGCGCGCTCTGTGATCAGCGCGGCGCATACCGGTACCCGACGTCAGGTTCACT 758
DB 645 CTCTCTCTCTCTCAATTGACAGCGCATCAGGCTCACCCGAGACGAAAGTGAACGT 704
QY 759 TTGACAGAGGAAACGAGATCATGTTGCGCGAGAGCCATCGCTCTTGTGCT 818
DB 705 CGTTCACAGAGGAGAGAGATCTGTAGCCCGGAGGCAATGGCGTCTTCAACT 764
QY 819 CGAGCACTGCTCTCGCGCGGAGAGGCTTCTGCTCAACGAAACGCGCTCTC 878

DB 765 CGAGCCCTCTGCTCGGCGCGAAGAGGCTCTGCTCAACGAGCCGCGTCTC 824
QY 879 CCGCTGAGAGGAGACCTCACTGTCAGCAATGCTCTGCTCTCGAGG 938
DB 882 ACATGATGAGCCACCTCGCTGTCAGAGACATGCTCTGCTCTCGAGT 884
QY 939 CTGACGCTTCAAGGTGAGGCTATGTCGCGACAGAGGCTCTGCGCGTTCAT 998
DB 885 GCTCAGGCAATGACGGTGAAGGATGTCGAGCGCGCTCTTCCACCTTCT 944
QY 999 CCACAGCTGCGCGCGCGACCCCGGCAAGTGAAGTGGCGGCAATCCGACGCT 1058
DB 945 TCAGACGTCAAGGCTCTCAACCGACGAGATGAGTGGGAAATCCGAACT 1004
QY 1059 CTTTCCGCTGCTGCTTGGCTTGAAGCAAGAGAGATCAAGTCAAGAGCAGCA 1118
DB 1005 CTTGAGGAGAGCGCTTGTGCTCCACATGAGAGAGTCAAGTCAAGAGCAGCA 1064
QY 1119 GGGCATTTTGGCAGAGACGCTACCGCTCCGACAGTCCCTCAATCTCTGCGCGCT 1178
DB 1065 GGGATTTCTCGCAGAGACGCTACCTTGGCAGAGTCTCTGAGTGGCTGCGCGCT 1124
QY 1179 CGTGAGAGCATGATGACGCTTACTGACTCTCTGCTGAGAAACA---CGACGAC 1235
DB 1125 CGTCAGGACCTCATTCAGCGCCACGCGCGCTCACCATGAGCGCGGCAATCGACGAC 1184
QY 1236 GCAACACCGCTCTCTGACCTGCAAGAAACGACGACCGGACGCTTCCAGGCT 1295
DB 1185 CGACACCTTCTTCAACGCTGAGAAACAAGCTTGCACACAGCGGCGCAATTTCCAGGCT 1244
QY 1296 GTGCGCTGTCTGATTTGATGAGAGAAACAGAGCTCGACCTGCGCTCATCGGCAAGCT 1355
DB 1245 TGCGGCTGTGCGCAACATGAGAGAACTGCGCTCGGAGCTGCGCAATCGGCAAGCT 1304
QY 1356 CACTTCAACGAGTGCACCGAGTTGCTCAAGCTGCTCAAGACGCGGCTGCTGCT 1415
DB 1305 CACTTCAACGAGTGCACCGAGTTGCTCAAGCTGCTCAAGACGCGGCTGCTGCT 1364
QY 1416 CCGCTGCGAGAGACCGTGTGCTCAACTCAAGCGCAAGGCTTGAATTCATCAATGCG 1475
DB 1365 CTTGCGCGGAGAACCTTCTGCTCTCTCACTGAGAGGCTGAGATGCGGCTGCT 1424
QY 1476 TGCTTACGCTTGGAGCTGCGCACTTGCACACCGGCTCATTAACCTTGTGCGAGCGC 1535
DB 1425 GCGCTTACACCTCGAGTTGGAGACCTGCGCAACCTGAGAGACGATGTCAGGCGCG 1484
QY 1536 AGAATGAGCAACAGCGCTCACTGCTGCTCTCATCTCGGCGCGGACATGCGCA 1595
DB 1485 TGAATGAGCAACAGCGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1544
QY 1596 GCGCCACGAGCTCTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
DB 1545 GTCCAAAGAGTCTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
QY 1656 GAGCTCCGCGCGATGAGTGAATCTTCAAGAGAGCTTCAAGAGCTTCTTCCAGCTCT 1715
DB 1605 CGACTTCCGCGCGAGAGTGAATCTTCAAGAGAGCTTCAAGAGCTTCTTCCAGCTCT 1664
QY 1716 CTTGAGAGACCTCGGACGAGTGAATCTTCAAGAGAGCTTCAAGAGCTTCTTCCAGCTCT 1763
DB 1665 CATGAGACACATTTTGTGCTGCGCATGACCGGCTGAGACCTGCGGACGAGCTGCTGA 1724
QY 1764 CGAGTCAAGAGAGCTTCAAGAGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
DB 1725 GAAAGTGAAG 1784
QY 1824 GCGCTGAG 1883
DB 1785 GCGCTGAG 1844
QY 1884 GCGCTTGCAG 1943
DB 1845 GTC-----GCTCTGCTGCGCGCGTCAACGCTGAGAGAGTGGCGCGCGAGAGTGC 1895

165 GACCGACTCGACGCTCGAAGCTCGAAGCTCTGCTCAACCTTCGAGAGAGCTGCTCTCGGC 224
219 CGCCCGGAGAGGGGGGAGAGGCGCGGCTCGAAGAGAGAGATCGCGACGCGCTCGA 338
225 CGCAGAGAGAGGGGAGGCGCTGCTCGCTCAAGAGAGAGAGATCGCTCAAGAGATTGA 284
339 CAAGAGCTGCACTTCTCTCAAGGCGCGAAGCTTCAGAACTCGGCTTCAGAGATCAACGCG 398
285 CAATTCGGTTCAGATTCTTGGCGCTCGCAACTCTCATAGAGCGCTTACGCGGCTCAAGACTGG 344
399 TTTGGTGGCTCGGCGCGAGAGAGAGATGAGAGATGAGAGTCAAGCTCCAGAAAGCGCTCAT 458
345 ATTGGCGAGATCCGAGACACCGCACCGAGAGAGAGAGATCGCTTCAGAGAGCTCTCTCT 404
459 CGAGACGAGAGCTTCGCGGCGCTGAGCGGACGAGCTCGCTCTCTTCAAGCTTCGAGCGCG 518
405 CGAGAGACGAGCTTCGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
519 CCTCGAGAGACGCTTCGCTCGCTGAGAGTCTCGCGGCGCGAATGATCTCGGCTCAACTC 578
465 TCTCGAGAACTCGCTTCCCTCGAGGTTGTTGCGGCGCGCATGACATTCGCGCTCAACAG 524
579 GCTCAGCGGCTCGCACTCGGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 638
555 CTGAGCGCGGCGCACTCGGCTGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 584
639 GAACACGCGCATCAAGCGCATCTGCTCCCTCGCGGCTCGCATCTCGGCTCGGCGCACT 698
585 CAACACGCGCATCAACCGCATCTGCTCCCTCGCGGCGCATCTCTGCTCGGCGCACT 644
699 CAGCGCGCTCTGCTCATCTCGCGGCGCGCATCAAGCTCAACCGCATCAAGCTCAAGCT 758
645 CT 704
759 TTTGACGAGAGAGAGAGAGATCATGTTGCGGCGAGAGCATCTCTCTCTCTCTCTCTCT 818
705 CCGTCAAGAGAGAGAGAGAGATCTCTGATCGCCCGAGAGAGAGATGAGCGCTCTCTCAACT 764
819 CGAGGAGAGTCTCTCTCGGCGCGCGAGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 878
765 CGAGCGCGCT 824
879 CGCTCTGATGAG 938
825 AGAGTATGAG 884
939 CTGAGCGCTCTCAAGTATGAG 998
885 GCTCAGCGGCT 944
999 CCAAGAGAGTCT 1058
945 TCAAGAGAGTCT 1004
1059 CTTTTCT 1118
1005 CTTGAG 1064
1119 GGGCATTTCT 1178
1065 GGGGATTTCT 1124
1179 CTTGAGAGAGATATATGAG 1235
1125 CTTGAGAGAGATATATGAG 1184
1236 CCAAG 1295
1185 CGAG 1244
1296 GTGCGCT 1355
1245 TGCGGCTGTGTGAG 1304

QY 1356 CAATTGAG 1415
Db 1305 CAATTGAG 1364
QY 1416 CTTGAG 1475
Db 1365 CTTGAG 1424
QY 1476 TGTGAG 1535
Db 1425 GCGTCAACCTCGAG 1484
QY 1536 AGAGATGAG 1595
1485 TGAAGTGGAG 1544
QY 1596 GAGCAG 1655
Db 1545 GTCCAG 1604
QY 1656 GAGCTTCGCGGAG 1715
Db 1605 GAGCTTCGCGGAG 1664
QY 1716 CTTGAG 1763
Db 1665 CATGAG 1724
QY 1764 CAGAGTCAAG 1823
Db 1725 GAGAGTCAAG 1784
QY 1824 GCGCTGAG 1883
Db 1785 GCGCTGAG 1844
QY 1884 GCGCTGAG 1943
Db 1845 GTC-----GCTCTGCTCGCGCGCGCTCAAGCTGAGAGAGAGAGAGAGAGAGAGAG 1895
QY 1944 GCGCATTCGCTCAAG 2003
Db 1896 GCGCATTCGCTCAAG 1955
QY 2004 GCGCGAG 2063
Db 1956 GCGCGAG 2015
QY 2124 GAGCAG 2183
Db 2076 CTTGAG 2135
QY 2184 CAAGATGCTCGGTA 2198
Db 2136 CAAGATGCTCGGTA 2150

Search completed: September 11, 2004, 14:50:22
Job time : 1351.06 secs

Dh 301 CGCGTCCAGAACGACGAGATCCGCGCACGCGTTCGACAGAGCGTGCATCTTCTCAAG 360
Qy 361 GCCAGCTTCAGAACTCGGTCTACGAGATCACACAGGCTTTGGTGGCTCGGCGACAG 420
Dh 361 GCCAGCTTCAGAACTCGGTCTACGAGATCACACAGGCTTTGGTGGCTCGGCGACAG 420
Qy 421 AGGACTGAGAGATCAGTCAAGCTTCAGAAAGCGCTTCATGAGCACACAGCTTCGCGCTG 480
Dh 421 AGGACTGAGAGATCAGTCAAGCTTCAGAAAGCGCTTCATGAGCACACAGCTTCGCGCTG 480
Qy 481 ACGCGAGCTCCGCTCGCTTCAGGCTCGAGCGCGCTTCGAGAAACAGCTTCGCTC 540
Dh 481 ACGCGAGCTCCGCTCGCTTCAGGCTCGAGCGCGCTTCGAGAAACAGCTTCGCTC 540
Qy 541 GAGGTGCTCCGCGCGCTTCAGTGCATCCGCTCAACTCGCTCAAGCTTCGCTGAGTGC 600
Dh 541 GAGGTGCTCCGCGCGCTTCAGTGCATCCGCTCAACTCGCTCAAGCTTCGCTGAGTGC 600
Qy 601 GTCGCGCTCGTCTCTTCAGAGCGCTTCACCACTTCCTGAAACACCGCTATCAGCCATC 660
Dh 601 GTCGCGCTCGTCTCTTCAGAGCGCTTCACCACTTCCTGAAACACCGCTATCAGCCATC 660
Qy 661 GTCGCGCTCGTCTCTTCAGAGCGCTTCACCACTTCCTGAAACACCGCTATCAGCCATC 720
Dh 661 GTCGCGCTCGTCTCTTCAGAGCGCTTCACCACTTCCTGAAACACCGCTATCAGCCATC 720
Qy 721 GCGCGCATACCGGCTCACCGCGCTCAAGGTTCAAGTTTTCAGAGAGAACGAGAG 780
Dh 721 GCGCGCATACCGGCTCACCGCGCTCAAGGTTCAAGTTTTCAGAGAGAACGAGAG 780
Qy 781 ATGATGTTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Dh 781 ATGATGTTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Qy 841 AAGAGAGTCTCGTCTGAGTCAACGAGAACGCGCGCTTCGCGCTGATGAGGAGGAG 900
Dh 841 AAGAGAGTCTCGTCTGAGTCAACGAGAACGCGCGCTTCGCGCTGATGAGGAGGAG 900
Qy 901 CTGCAAGATTCGCAATGCTTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Dh 901 CTGCAAGATTCGCAATGCTTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Qy 961 GCGATGCTCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Dh 961 GCGATGCTCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Qy 1021 CCCGCGCAGGTCGAGGTCGCGCGAGCAATCCGCAAGCTTCCTTCGCGCTGCTGCTGCTC 1080
Dh 1021 CCCGCGCAGGTCGAGGTCGCGCGAGCAATCCGCAAGCTTCCTTCGCGCTGCTGCTGCTC 1080
Qy 1081 GTTGAAGCAGAGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Dh 1081 GTTGAAGCAGAGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1141 TACCGCTTCGCGAGCTGCTCAAGTTCCTCGGCGCGCTTCGCTGAGAGAGATGAGAGAG 1200
Dh 1141 TACCGCTTCGCGAGCTGCTCAAGTTCCTCGGCGCGCTTCGCTGAGAGAGATGAGAGAG 1200
Qy 1201 TACTGAGATCTTCGCTTCGAGAAACAACAACAACAACAACAACAACAACAACAACAACA 1260
Dh 1201 TACTGAGATCTTCGCTTCGAGAAACAACAACAACAACAACAACAACAACAACAACAACA 1260
Qy 1261 AACAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1320
Dh 1261 AACAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1320
Qy 1321 AAGACAGAGCTTCGAGCTTCGCTTCATGAGAGAGTCAAGTTCAGAGAGAGAGAGAGAG 1380
Dh 1321 AAGACAGAGCTTCGAGCTTCGCTTCATGAGAGAGTCAAGTTCAGAGAGAGAGAGAGAG 1380
Qy 1381 CTCAAGCTTCGAG 1440
Dh 1381 CTCAAGCTTCGAG 1440

Qy 1441 AACTATCACCGGAGAGGCTTCGAGCAATTCGAGTTCGCTTCGAGAGCTTCGAGCAG 1500
Dh 1441 AACTATCACCGGAGAGGCTTCGAGCAATTCGAGTTCGCTTCGAGAGCTTCGAGCAG 1500
Qy 1501 CTTCGCAACCGGCTTCAGTTCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Dh 1501 CTTCGCAACCGGCTTCAGTTCGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Qy 1561 TCGCTCGCTTCATATTCGCGCGCGCGCGAGCTTCGAGAGAGAGAGAGAGAGAGAGAG 1620
Dh 1561 TCGCTCGCTTCATATTCGCGCGCGCGCGAGCTTCGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1621 CTGCGCTTCGAG 1680
Dh 1621 CTGCGCTTCGAG 1680
Qy 1681 TTCAAGAGAGAGTTCAG 1740
Dh 1681 TTCAAGAGAGAGTTCAG 1740
Qy 1741 CTGAGAGTTCAG 1800
Dh 1741 CTGAGAGTTCAG 1800
Qy 1801 ACGAGAGAGTTCAG 1860
Dh 1801 ACGAGAGAGTTCAG 1860
Qy 1861 GTGCTGAGAGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1920
Dh 1861 GTGCTGAGAGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1920
Qy 1921 TGAAGAGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1980
Dh 1921 TGAAGAGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 1980
Qy 1981 TGGAG 2040
Dh 1981 TGGAG 2040
Qy 2041 CTGTAAGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2100
Dh 2041 CTGTAAGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2100
Qy 2101 GCGCTGAG 2160
Dh 2101 GCGCTGAG 2160
Qy 2161 GCGCGCATCAACAGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2220
Dh 2161 GCGCGCATCAACAGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2220
Qy 2221 GAGCGCGCTTCAG 2280
Dh 2221 GAGCGCGCTTCAG 2280
Qy 2281 CTTCATACAGAGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2340
Dh 2281 CTTCATACAGAGTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2340
Qy 2341 CCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2400
Dh 2341 CCGCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 2400
Qy 2401 AGTTCAAAAAAAAAAAAAA 2419
Dh 2401 AGTTCAAAAAAAAAAAAAA 2419

Patent No. 6355468
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
INVENTOR: Koester, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
TITLE OF INVENTION: Using Same
FILE REFERENCE: 29479/500NSC
CURRENT APPLICATION NUMBER: US/09/624,693A
CURRENT FILING DATE: 2000-07-24
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
LENGTH: 2475
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc_difference
LOCATION: (13_34, 46, 49, 51, 57, 59, 68, 69, 73, 75 - 77, 79, 82, 84,
OTHER INFORMATION: n = A or C or G or T; "n" indicates no consensus at that position
OTHER INFORMATION: Description of Artificial Sequence: Consensus
OTHER INFORMATION: Sequence of SEQ ID NOS: 12, 16, and 18
Patent No. 6355468
US-09-624-693A-20

Query Match 57.3%; Score 1629; DB 4; Length 2475;
Best Local Similarity 78.0%; Pred. No. 7,3e-307;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;
37 ATGCGCCCTTCCTTGAAGTCTGCTGCGCAGCAGCTGCGCAACGAGCTTACCAACGCGTGG 96
1 ATGCGCCCTTCCTTGAAGTCTGCTGCGCAGCAGCTGCGCAACGAGCTTACCAACGAGTNG 60
97 CAGCGCGCTCGACCAAGTCTGCTGCGCGCCCGCCTTGGCTTCCCGCGCAGCGC -CG 154
61 CAGCGCGNNCCGNN 120
155 GCCTGATGAGCGACCGCGCGACAGTGCAGCTGAGATGATGATGATGATGATGATGATGAT 214
121 GCTCGCTCTCCGACACCGCAGNNNAGCAGCTGAGATGATGATGATGATGATGATGATGAT 180
215 ACCCGACCGAC---GACGTCTGAGCTCAGCGGGTACAGCTTACCGCTGCGTGG 271
181 ACCCGACCGNNNAGCAGNN 240
272 TCG 330
241 TCG 300
331 CGCGTGAAGAAGAGTCTGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
301 AAGATCGAAGAAAGTCTGAGTCTTCTCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360
391 ACCCGCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 450
361 ACAGCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
451 GCGCTATGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 510
421 GCGCTCTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
511 GAGCGCGCGCTCGAGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 570
481 GAGCGCGCGCTCGAGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
571 GTCACTCTGCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 630
541 GTCACTCTGCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
631 AACTCTGTAACACCGCATCAAGCCCATGTCCTCCCTCGCGGCTCATCTCGCGCTCG 690
601 AACTCTCTCAACACCGCATCAAGCCCATGTCCTCCCTCGCGGACCATCTCGCGCTCG 660

QY 691 GCGCACTTCAAGCCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 750
DB 661 GCGCACTTCTCCCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 751 GTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 810
DB 721 GTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
QY 811 TTTGCTCTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
DB 781 TTTGCTCTGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 871 GCGCTCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
DB 841 GCGCTCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 931 TCGAGAGCTTGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 990
DB 901 TCGAGAGAGCTTGAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 991 CCGTTCATCGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1050
DB 961 CCGTTCATCGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1051 CGCAGCGCTCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1110
DB 1021 CGCAGCGCTCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 1111 GACGACGAGGAGCTTCTTCCGAGAGACCGCTTCCGAGAGTGGCTGCTGCTGCTGCTGCT 1170
DB 1081 GACGACGAGGAGCTTCTTCCGAGAGACCGCTTCCGAGAGTGGCTGCTGCTGCTGCTGCT 1140
QY 1171 GAGCGCTCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1227
DB 1141 GAGCGCTCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1228 ACAGCAGCTGACCAACCGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1287
DB 1201 TCGAGAGCTGACCAACCGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1288 TTTCAAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1347
DB 1261 TTTCAAGGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1348 GCGAAGCTCAACTTCAAGCAGTGAACGAGTGTCTCAACGCTGCTCAACCGCTGCTGCT 1407
DB 1321 GCGAAGCTCAACTTCAAGCAGTGAACGAGTGTCTCAACGCTGCTCAACCGCTGCTGCT 1380
QY 1408 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
DB 1381 CTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
QY 1468 CACTGCTGCTTACGCTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
DB 1441 GCGCGGCGCTTACGCTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1528 CAGCGCGAGAGTGGGCAACGAGGCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1587
DB 1501 CAGCGCGAGAGTGGGCAACGAGGCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1588 ACTGCGAGGCGCAAGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
DB 1561 ACTGCGAGGCGCAAGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY 1648 CAGGCGCTGAGCTTCCGCGAGTGAAGCTGAGCTTCAAGAAAGAGTTCAGCCGCTTCTC 1707
DB 1621 CAGGCGCTGAGCTTCCGCGAGTGAAGCTTCAAGAAAGAGTTCAGCCGCTTCTC 1680
QY 1708 CCGACTCTCTCAAGAGAGCTTCCGCGAGTTCAGCCGCTTCAAGAGTTCAGCCGCTTCTC 1755
DB 1681 CCGACTCTCTCAAGAGAGCTTCCGCGAGTTCAGCCGCTTCAAGAGTTCAGCCGCTTCTC 1740
QY 1756 CTTGCGCTGAGAGTCAAGAGGCTCAAGAGGCTCAAGAGGCTCAAGAGGCTCAAGAGGCT 1815

Dp	1741	CTCGNAGCAAAAGGTCACAAAGNCGCTTCACAAAGCGNCTCGAGCGACCACTCTGTAACAC	1800
Qy	1816	CTGAGAGCGGCGCTGCGACACGACCGCTTTCTCGTAACGGACCGGACCGTCTCGACGCTCTC	1875
Dp	1801	CTGAGCGCGGCTGCGACACGACCGCTTCTCGTTTCGACCGGACCGTCTCGACGANNNNN	1860
Qy	1876	TCGTCCTCGCCCTCTGCCA--ACGTCACCCCTTATCGCGCGTCAACGGGTGAAAGTTGCTC	1933
Dp	1861	NNGTCTCGCGNNNNGCCANNAGGTCCTGCTCGCGCCGTCACACGCTGTGAAAGTGGCT	1920
Qy	1934	CGGCGGAGAGGCTCATCTGCTCATCGCGCGAGGTGCGCAACCGCTTCTGCGACAGCCGT	1993
Dp	1921	CGCGCGAGAAAGGCTCATCTGCTCATCGCGCGANGTCCCGACGNCCTTCTGACGCGCCGT	1980
Qy	1994	CTTCGAGAGGCGCGGCGACGACGATCATCTCGCGCGGACCGCGCGTCTGTAACGTTGCG	2053
Dp	1981	CGTGTGTGTGCGCCCGGCGTCGNGATCATCTGCGCGGACGCGGCTCTGTACTGTTCG	2040
Qy	2054	TGCGCGAGAGCTCGGCTGTGACGGCGCGCGCGGCGACGTTTGTGCGCGTGCAGCAGG	2113
Dp	2041	TCGCGAGAGGCTCGGCGTCAAGGCGCGCGCGCGACGCTTCTCGGCAAGCGAGAGG	2100
Qy	2114	AGACGATCGGAGACCAAGCTTCGCGCATCTACGAGGCGTCAAAGACCGCGCGATCAAC	2173
Dp	2101	TGACGATCGGACCAACGCTCTCCGCACTTACGAGGCGTCAAAGNCGCGCATCAAC	2160
Qy	2174	ACGTCCTCTGTCAAGTGTCTCGGTA-AGGCCGAGCAAGCTTCGCGCTAGACGCGCGCTC	2232
Dp	2161	ACGTCTCTGTCAAGTGTCTCGGTAAGNNCNCNNNNCNCNNCTCGCTNNNNNCNNNCNC	2220
Qy	2233	ACCCCAAGACCAAGCTTTTGAAGTGTGTCGCGCAAGAACGGAATTCTCTCATACACA	2292
Dp	2221	NNNCNNNNNNNNCTTTNGNNNTCGNNTCTTACACNNNNCGGAAATNNCNCNNNNNN	2280
Qy	2293	TGTGCGCTTACTCTCTCGCCGTCATCAAGTCTCTCAGTTCTTTCGTATCCGCGCTCTC	2352
Dp	2281	TNNNNCTTNNCTNNCTCNCNNNNNANCGNCTNNNNNNCTNNNNNNCTNNNNNNNNCNC	2340
Qy	2353	GGTCGTCACGTGATAGAGCGTGAATGATTCAGAGTCTTGGAATTCAGAAAAA	2411
Dp	2341	NGNNNNCANNNNCNCNTTNNNNNANCGANNGATNNANAGNTTNCGNANNCNNNNN	2400
Qy	2412	AAAAAAAA 2419	
Dp	2401	NNNAAAAA 2408	

```

1 RESULT 3
2 US-09-627-216A-9
3 : Sequence 9, Application US/09627216A
4 : Patent No. 6368837
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: Sariaslani, Sima F
8 : APPLICANT: Tang, Xiao-Song
9 : APPLICANT: Qi, Wei Wei
10 : APPLICANT: Vannelli, Todd
11 : APPLICANT: Gatenby, Anthony
12 : TITLE OF INVENTION: Bio-production of para-Hydroxycinnamic Acid
13 : FILE REFERENCE: BC1009 US NA
14 :
15 : CURRENT APPLICATION NUMBER: US/09/627,216A
16 :
17 : PRIOR FILING DATE: 2000-07-27
18 : PRIOR APPLICATION NUMBER: 60/1147,719
19 : PRIOR FILING DATE: 1999-08-06
20 : NUMBER OF SEQ ID NOS: 14
21 :
22 : SOFTWARE: Microsoft Office 97
23 :
24 : SEQ ID NO 9
25 :
26 : LENGTH: 2151
27 :
28 : TYPE: DNA
29 :
30 : ORGANISM: Artificial Sequence
31 :
32 : FEATURE:
33 :
34 : OTHER INFORMATION: Description of Artificial Sequence: mutant from Rhodocortula glutinis

```

US-09-627-216A-9

Query Match 49.5%; Score 1196.6; DB 4; Length 2151;
Best Local Similarity 74.0%; Pred. No. 3.8e-223;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3

QY	99	CGCGCTCCACCAAAAGTCGGCTGCGGGGCCCACTTCGGGTCTCGCGCGACCGCCGGGCT	158
Db	45	CGTGCATCCGCAAAAGAGAGCTCTCATATGGCCCTGCACCACTTCGATCGAGCTC	104
QY	159	CGATGGCCAGCGCGCGACCAAGTCGAGCTCGAGATCGTGGAGAGCTCTCAACGACC	218
Db	105	GCACCTGCCCAACCAACAGAGTCACGAGAGTGCATCTCGAAGAAATGCTCGCCGCC	154
QY	219	CACCGAGAGCTGCTGAGTCTAGCGGGTAACGCTTACCGTCCGTACGTTGTGCGCG	278
Db	165	GACCGACTCCACGCTGAACTGACGGCTACTCGCTCAACTCGGAGACGTCGTCGGC	224
QY	279	CGCCCGCAAGGGGCGAGGGTTCGGCTCCAGAACGACGACGAGATCCGCGCACGCTCGA	338
Db	225	CGCAGAGAAAGGACAGGCTGTCCGGCTCAAGACAGCGACGAGATCCGCTCAAAAGATGA	284
QY	339	CAAGAAGGTGAGCTTCTCTTAAGGCCAGTTCAGAACTCGGTCTACGAGTCAACAGG	398
Db	285	CAATTCGGTTCGAGTCTTTCGGCTCGCACTCCATGAGCTCTTACGAGCTCAAGTGG	344
QY	399	TTTCGTGCTCGGCGCGACACGAGGACTGAGAGTACATGACGCTTCAGAAAGCGCTCAT	458
Db	345	ATTTCGCGATCCGACGACACCGGACCGAGGACGCATCTCGTCCAGAAAGCTCTCT	404
QY	459	CGAGCAACAGCTTGGGGGTGACGCCGACGTCGCTCTGCTCTTACAGGTCCGACGGG	518
Db	405	CGAGCACACACTTGGGAGTTCCTCCCTTCGTCGTCGATCGTTCGCTTCGGCCCGCG	464
QY	519	CCTCGAACAACGCTTCGCTCGAGGTGATCCGCGCGCCATGATCATTCGCGCTCAATC	578
Db	465	TCTCGAAGAACTGGCTTCCCTCGAGTGTGTCGCGGCCCATGACAAATCCGCTCAACAG	524
QY	579	GCTACGCGTGGCCACTCGGCCGCTCGCTCTGTGCTCTTGAAGCGCTCAAACTCTT	638
Db	525	CTTGACCCCGGCGCACTCGGCTGTCCGCTGTGTCTCGAAGCGCTCAACCACTTCT	584
QY	639	GAACCAACCGCATACGCGCATGTCGCCCTCCGCGGCTCCATCTCGGCGTCGCGGCACT	698
Db	585	CAACACAGGATACCCCATGTCTCCCTCCGGGACCATCTCTGTCGCGGCACT	644
QY	699	CACCCGCTCTGTATATGCGCGGCGCATACCGGTACCCGACGTCAGATTCAGT	758
Db	645	CTCTCTCTCTCTATGATTCGACGCGCATACGGGTCAACCGAAGCAAGGTGCAGCT	704
QY	759	TTTGACGAGGAAACCGAAGATCATTTTGCGCGGAGGCAATCGCTCTTTGAT	818
Db	705	CGTCCACGAGGCGAAGAAAGATCTCTATAGCCCGGAGGCAATGCGCTCTTCAACT	764
QY	819	CGAGGAGTCTGCTTCGGCCCGAAGAGAGGTCTGATCTGTCTCAACGAAACGCGCTCTC	878
Db	765	CGAGCCCGTCTCTTCGGCCCGAAGAGAGGTCTGATCTGTCTCAACGAAACGCGCTCTC	824
QY	879	CGGCTGATGAGGACCTTCAGTTCGACGATCGCATGCTTCGCTCTCTCGAGGC	938
Db	825	AGATGAGATGSCAACCTTCGCTCTGACGACGACAAATCTTCGCTCTTCGCACTC	884
QY	939	CTTGAAGGCTCTCAAGGTGAGGCGCATGTGCGGCGACAGAGGCTGTTGCGCGCTCAT	998
Db	885	GCTCAAGGCGCAAGAGGTGAAAGATGTCGCGCACGCGGCTCTGTTCAACCTTCT	944
QY	999	CCACAGCTCTGCGCCCGCACCCCGGCGCAGTTCGAGATCGCGCGCAATCCGACGT	1058
Db	945	TACGACGTCAAGCGCCCTCACCCGACGCAAGATCGAATGCGGGAAATCCGCAACT	1008
QY	1059	CCTTTCGCGCTGTGTTGCGGTGACACGAGAGAGGTCAAGTCAAGACGACGA	1118
Db	1005	CCTCAGAGAAACCGCTTGTGCTTCCACATGAGAGAGGTCTAGTCAAGACGACGA	1068

QY 1119 GGGCATTTCTGCGAGAGACCGCTACCGCTCCGACGCTCGCTCACTTCTCGGCCGCT 1178
 DB 1065 GGGCATTTCTGCGAGAGACCGCTACCGCTCCGACGCTCGCTCACTTCTCGGCCGCT 1124
 QY 1179 CGTGGAGACATGATGACGCGCTTACTGACTCTCTCGCTGGAACAACA---CGACGAC 1235
 DB 1125 CGTGGAGACATGATGACGCGCTTACTGACTCTCTCGCTGGAACAACA---CGACGAC 1184
 QY 1236 CGACACCGCGCTCCGCTCGAGCTCGAAGAACAGACGCGGACGCGGCACTTCCAGGC 1295
 DB 1185 CGACACCGCGCTCCGCTCGAGCTCGAAGAACAGACGCGGACGCGGCACTTCCAGGC 1244
 QY 1296 GTCCGCTGTGGGCAACACCATGAGAGACTCGCTCGGCTCGCCGAGATCGCAAGCT 1304
 DB 1245 GTCCGCTGTGGGCAACACCATGAGAGACTCGCTCGGCTCGCCGAGATCGCAAGCT 1304
 QY 1356 CAACCTTCACGAGCTGACCGAGCTGCTCAACGCGCTGCAATGAACGCGGCTGCTGCTG 1415
 DB 1305 CAACCTTCACGAGCTGACCGAGCTGCTCAACGCGCTGCAATGAACGCGGCTGCTGCTG 1364
 QY 1416 CCTCGCTGCGAGAGACCGCTCGCTCACTTACGCGCAAGGCTTGGACATTCATCGC 1475
 DB 1365 CCTCGCTGCGAGAGACCGCTCGCTCACTTACGCGCAAGGCTTGGACATTCATCGC 1424
 QY 1476 TCGTTACGTTGGAGCTCGGCTCGACCTTGGCAACCGCTGCACTTCTGCGAGCGCG 1535
 DB 1425 GCGGTACACCTTGAGTTGGAGCACCTGCGCAACCTTGTGAGAGATGTCAGCCGCTG 1484
 QY 1536 AGAGATGGGCAACCAAGGCGCTCACTCGCTGCTCACTTCTCGGCGCGCACTTGGCGA 1595
 DB 1485 TGAAGATGGGCAACCAAGGCGCTCACTCGCTGCTCACTTCTCGGCGCGCACTTGGCGA 1544
 QY 1596 GGGCAACGAGCTCTTCT 1555
 DB 1545 GTTCAACGAGCTCTTCT 1604
 QY 1656 CGACCTTCGCGGAGTGGAGCTGAGCTGAGCTGAGAGAGAGTTGAGACCGGCTTCCGAGCT 1715
 DB 1605 CGACCTTCGCGGAGTGGAGCTGAGCTGAGCTGAGAGAGAGTTGAGACCGGCTTCCGAGCT 1664
 QY 1716 CCTCGAGAGCACTTCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1763
 DB 1665 CATCGAGAGCACTTTCGCTCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1724
 QY 1764 CGAGGCTCAAGAGGCGCTCAACAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1823
 DB 1725 GAGAGTGAACAGAGCGCTGCGCAAGGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAG 1784
 QY 1824 GCGCTGAGAGAGAGCTTCTCTGTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1883
 DB 1785 GCGCTGAGAGAGAGCTTCTCTGTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1844
 QY 1884 GCGCTGAGAGAGAGCTTCTCTGTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1943
 DB 1845 GTC-----GCTCTGCTCGCGCGCTGCAACGCTGGAAGGCTGCGCGCGAGCTC 1895
 QY 1944 GAGCATCTGAGTGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2003
 DB 1896 GAGCATCTGAGTGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1955
 QY 2004 GCGGAGAGAGAGAGAGCTTCTCTGTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2063
 DB 1956 GCGGAGAGAGAGAGCTTCTCTGTAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2015
 QY 2064 GCTCGGAG 2123
 DB 2016 GCTTGGGAGTGAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2075
 QY 2124 GAGCAACGATCGAGATTCAG 2183
 DB 2076 CTGGAACGTTTCAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2135

QY 2184 CAAGATGCTGCGCTA 2198
 DB 2136 CAAGATGCTGCTTA 2150
 RESULT 4
 US-09-765-873A-9
 / Sequence 9, Application US/09765873A
 / Patent No. 6521748
 / GENERAL INFORMATION:
 / APPLICANT: Tang, Xiao-Song
 / TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
 / FILE REFERENCE: BCI009 US CIP
 / CURRENT APPLICATION NUMBER: US/09/765,873A
 / PRIOR FILING DATE: 2001-01-19
 / PRIOR APPLICATION NUMBER: US 09/627,216
 / PRIOR FILING DATE: 2000-07-27
 / NUMBER OF SEQ ID NOS: 38
 / SOFTWARE: Microsoft Office 97
 / SEQ ID NO 9
 / LENGTH: 2151
 / TYPE: DNA
 / ORGANISM: mutant from Rhodotorula glutinis
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (1)..(2151)
 US-09-765-873A-9
 Query Match 49.5%; Score 1196.6; DB 4; Length 2151;
 Best Local Similarity 74.0%; Pred. No. 3.8e-223;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;
 QY 99 CCGCCCTCGACCAAGTGGGCTGGGCGGCCCACTTGGGCTCTCCGCGAGCGGCGCT 158
 DB 45 CGTGCATCCGAAAGAGAGCTGCAATGCGCTCGACCACTTCCGATCGGAGCTC 104
 QY 159 CGATGGCAGCGCGCGGACAGTGCAGCTGAGATGATGAGAGAGCTCTCAGCGACC 218
 DB 105 GCACCTGCCAACACCCAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 164
 QY 219 CACCAAGAGCTGCTGAGCTCAGGAGTACAGCTCAGCTCAGCTCAGCTCAGCTCAG 278
 DB 165 GACCACTGAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 224
 QY 279 GCGCGCAAGGGGCGCAGAGTCCGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 338
 DB 225 CCGAG 284
 QY 339 CAAAGAGCTGACTTCTCAAGGCGCAGCTTCAAGACTCGCTTACAGAGTACACAGCG 398
 DB 285 CAATCGGTGAGTTCTTGGCGTCCCACTTCATGAGAGAGAGAGAGAGAGAGAGAGAG 344
 QY 399 TTTTGGTGGCTGGGCGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 458
 DB 345 ATTGGCGAGTCCGAG 404
 QY 459 CGAGCAGAGCTCTCGGCGGTGACCGAGAGTCCGCTCAAGCTTCAAGAGAGAGAGAG 518
 DB 405 CGAGCAGAGCTCTCGGCGGTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
 QY 519 CTTGAGAGAGAGCTTCCGCTGAGAGTGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAG 578
 DB 465 TCTCGAGAGAGCTGCTCCCTCGAGAGTGTCTCGGAGAGAGAGAGAGAGAGAGAGAG 524
 QY 579 GCTCAGAGTGGCAGCTGGGCGCTCGGCGCTGCTGCTTGAAGGAGTCAACACTTCTT 638
 DB 525 CTTGAG 584
 QY 639 GAACCAAG 698
 DB 585 CAACAG 644

QY 699 CAGCCGCTCTCGATACGCGCGCCATCACCGCTGACCCCGAGCTCAAGCTTCAAGT 758
Db 645 CT 704
QY 759 TTGGACAGGAGAACCGAGAAAGATGATGTTGGCGCGAGGCACTCTGCTTTTGGTCT 818
Db 705 GGTCCAGAGGAGGCAAGGAGAGATCTGTACGCGCGCGAGGAGATGGGCTTTTCAACT 764
QY 819 CGAGGAGATGTCCTCGCGCCGAGAGAGGATCTCGCTCAAGCAAGCAAGCGCGCTC 878
Db 765 CGAGCCGCTCTGCTCTGGGCCGAGAGAGGCTCTCGCTCTCTCTCTCTCTCTCTCT 824
QY 879 CGCTCTGATGGCAACCTCTCATGTCGACATCTGCAATGCTCTCTCTCTCTCTCT 938
Db 825 AGCATGATGGCAACCTCTCATGTCGACATGCTCTCTCTCTCTCTCTCTCTCTCT 884
QY 939 CTGACGAGCTCTCAGCGTGGAGGCAATGCTGGCGCAGAGAGGCTCTCTCTCTCTCT 998
Db 885 GCTCAGGCGCATACCGCTGAGAGCATGATGTCGAGCAAGCGCTCTCTCTCTCTCTCT 944
QY 999 CACGACGTCCTGCGCGCCGACCGCGCGAGTCTGAGATCTGCGCGCAACATCTCGCAGCT 1058
Db 945 TCACGACGTCACGCGCCCTCACCCGACGATGCAAGTCGCGGGGAAATCTCCGCAAGCT 1004
QY 1059 CCTTCCGCTGCTGCTCTGCGCTTGGAGACGAGAGAGATGCAAGGTCAGAGACGA 1118
Db 1005 CTTGAGGAGAAAGCGCTTGTGTCTGTCACCATAGAGAGAGATGCAAGGTCAGAGACGA 1064
QY 1119 GGGCATCTTCTGCGCAGACCGCTAACCGCTCGCAGCTCTGCTCTCTCTCTCTCT 1178
Db 1065 GGGCATCTTCTGCGCAGACCGCTAACCGCTCGCAGCTCTGCTCTCTCTCTCTCTCT 1124
QY 1179 CGTGGAGACATGATGACAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1235
Db 1125 CGTACGAGACCTCATCTACGCGCCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 1184
QY 1236 CGACAAACCGCT 1295
Db 1185 CGACAAACCT 1244
QY 1296 GTCCGCTGTCGATTTGATGAGAGAACCGAGTCTGCTCTCTCTCTCTCTCTCTCT 1355
Db 1245 TCCGCTGTGGCCAAACCATGAGAGATCTGCTGCGCTCTCTCTCTCTCTCTCTCT 1304
QY 1356 CAATCTCAAGCAGTGAACCGAGTTGCTCAAGCTGCTGCTGCTGCTGCTGCTGCTG 1415
Db 1305 CAATCTCAAGCAGTGAACCGAGTTGCTCAAGCTGCTGCTGCTGCTGCTGCTGCTG 1364
QY 1416 CCTCGCTGCGAGAGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1475
Db 1365 CCTCGCGCGGAAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1424
QY 1476 TGGCTTACGTTTGGAGCTCGGCACTTGGCCATCCCGTCACTTCTCTCTCTCTCT 1535
Db 1425 GGCCTACACTCTGGAGTTGGAGACCTGCGCAACCTGTGACGCGATGTCTCAAGCG 1484
QY 1536 AGAGATGGGCAACAGGCGCTCAACTCGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1595
Db 1485 TGAAGATGGGCAACAGGCGCTCAACTCGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1544
QY 1596 GGCCTACAGATCT 1655
Db 1545 GTCCAAAGACGTCCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1604
QY 1656 CGACCTCGCGGATGAGCTCGACTTCAAGAAACATGAGCCCGTTTCTCCGAGCTCT 1715
Db 1605 CGACTTGGCGCGAGCGAGTTGAGTTTCAAGAAACATGAGCCCGTCTCTCTCTCT 1664
QY 1716 CTTCCAGAGACCTCTGAGCTGAGCTCTCAAGTCA-----ACGCACTTGGCT 1763
Db 1665 CATGACAGGACCTTGGCTCTCGCATGACCGGCTGCAACCTGCGGAGAGAGCTCTCTGA 1724

QY 1764 CGAGTCAAGAGGCGCTCAACAGCGCTCTGAGACAGACGATACGACTGAGCC 1823
Db 1725 GAAGGTGAACAGAGCTCTGCGCAAGCGCTGAGAGACCACTGTACGACTCTGCTCC 1784
QY 1824 GCGCTGGCAGACGCTCTCTGTCGATGCGGACCGTCTGAGCTCTCTCTCTCTCT 1883
Db 1785 GCGCTGGCAGACGCT 1844
QY 1884 GCGCTCTGCAACGTCACCTTACTGCGCGTCAAGCGGTGAAGTTGCTCTCTCTCT 1943
Db 1845 GTC-----GCTCTGCTCTGCGCGCTGAGAGCTGAGAGTCTCTCTCTCTCTCT 1895
QY 1944 GGCATCTTCTGCTCAGCGCGAGGTCGCAACCGCTTCTGAGACGCGCTTTTTCAGAGC 2003
Db 1896 GGCATCTTCTGCTCAGCGCGAGGTCGCAACCGCTTCTGAGAGTCTCTCTCTCTCT 1955
QY 2004 GCGCGCGACAGATCT 2063
Db 1956 GCGCGCGCTCTGATCT 2015
QY 2064 GCTCGCGCTGAGAGGCGCGCGCGAGCGATGTTGTCTGCGGTGAGAGAGAGATCGG 2123
Db 2016 GCTTGGCTTCAAGGCGCGCGCGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2075
QY 2124 GAGCAAGCTCTGCGCATCTTGAAGCATCAAGACGCGCGCATCAACGCTCTCTCT 2183
Db 2076 CTGCAACGCTTCAAGATCTTACAGAGCATCAAGTCTGCGGAGATCAACACGCTCTCT 2135
QY 2184 CAAGATCTCTCGCTA 2196
Db 2136 CAAGATCTCTCTTA 2150

RESULT 5
US-09-624-693A-18
; Sequence 18, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocistara, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/50NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Rhodotorula toruloides
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ..(2148)
US-09-624-693A-18

Query Match 49.5%; Score 1196.6; DB 4; Length 2439;
Best Local Similarity 74.0%; Pred. No. 3; 9e-223;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CGCGCTCCAGCAAGCTGCGGCGCGCCCACTTCTCTCTCTCTCTCTCTCTCTCTCT 158
Db 45 CGTGGATCTCGCAAGAGGCTGTCAATGAGCGCTGACCAACTCTGCAAGTCTGCGAGCTC 104
QY 159 CGATGGCAGCGCGCGCGCAACGATCTGAGATCTGCGAGAGCTCTTCAAGGACCC 218
Db 105 GCACTTGGCCCAACCAACGAGTCAACGAGTCAATCTGTGAGAGAGTCTCTGCGCGCC 164
QY 219 CACGACAGAGTGTGAGCTAGCGGATACAGCTCAACCGCTCTGAGAGTGTGAGGCGC 278
Db 165 GACGAGTCAAGCTGAGTCAAGCTGAGCGCTACTCTGCTCAACTCTGAGAGAGTGTCTGCGC 224

279 CGCCGCAAGGGGCGAGGCTCCGCTCAGAAAGAGAGATCCGCGACGCTCGA 338
225 CGCAGAGAGGGGCGGCTGTCCGCTCAAGAGAGAGAGATCCGCTCAAAAGTTGA 284
339 CAGAGGCGTCACTTCTCAAGGCGGCTTCAAGAACTCGGTCTTACGAGTCAACAGGG 398
285 CAATTCGGTCAAGTTCTTGGCGCTGCAACTCTCATAGCGGTCTACGGGCTCAGACTG 344
399 TTTCGGTGTCTCGGCGCAGACAGAGAGATGAGATGAGCTCAGGCTCAAGAGGCGCTCAT 458
345 ATTGGCGGATTCGAGACAGACCGGACCGAGAGAGGCTCATTCGCTTCAAGAGGCTTCT 404
459 CGAGCAGGCTGTCCGCGCTGAGCGGCGGCTCGCTCGCTTCAAGCGCTGAGCGG 518
405 CGAGGACGAGCTCGCGGTCTTCCCTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 464
519 CCTGAGAAACGCTTCCGCTGAGGCTGTCCGCGGCGGCTGAGTCACTCGGCTCAACTC 578
465 TCTCAGAACTCGCTTCCCTCGAGGTTGTTGCGGCGGCTGAGCAATTCGCTCAACAG 524
579 GCTCAGGCTGCGCTCGGCGCTCGGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
525 CTGACCGCGGCGCTCGGCTGTCCGCTGTCCGCTGTCTGCTGAGGCGGCTCAACACTTCT 584
639 GAACCAAGGATCAAGCGGCTGTCCCTTCCGCGCTTCACTTCCGCGCTGAGCGGCT 698
585 CAACCAAGGATCAAGCGGCTGTCCCTTCCGCGCTTCACTTCCGCGCTGAGCGGCT 644
699 CAGCGCGCTGTCAATCGCGGCGGCTGAGCGGCTCAAGCGGCTCAAGGCTCAAGT 758
645 CTCTCTCTCTCTCAATTGAGGCGGCTGAGGCTGAGCGGCTCAAGCGGCTCAAGT 704
759 TTTCAGCAGAGGAGCGAGAAATCATTTTTCGCGGAGGCTCATCTGCTCTTTGCTCT 818
705 CGTCAAGGAGGAGGAGAGAAATCTGTACGCGGAGGCGGCTGAGCGGCTCTTCAACT 764
819 CGAGGAGTGTCTCTCGGCGGCGGAGAGAGGCTGAGTGTCTGCTCAAGGAGGCGGCTCT 878
765 CGAGCGCGTGTCTCTCGGCGGAGAGAGGCTGAGTGTCTGCTCAAGGAGGCGGCTCT 824
879 CGCCTGAGTGGGAGACCTCACTGCTGAGAGCTGAGAGCTGCTGCTGCTGCTGAGG 938
825 AGCATGATGGGAGACCTCTGCTGCTGAGAGAGAGAGTCTGCTGCTGCTGCTGAGG 884
939 CTTCAGGCTCTCAAGGAGGAGGCTGAGGCTGAGGAGGCTGCTGCTGCTGCTGAGG 998
885 GCTCAAGGAGGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGG 944
999 CAGCAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1058
945 TACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1004
1059 CTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
1005 CTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1064
1119 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1178
1065 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1124
1179 CGTGAAGAGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1235
1125 CGTGAAGAGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1184
1185 CGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1244
1296 GTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1355
1245 TGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1304
1356 CAATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1415

1305 CAATTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1364
1416 CTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1475
1365 CTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1424
1476 TGCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1535
1425 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1484
1536 AGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1595
1485 TGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1544
1596 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1655
1545 GTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1604
1656 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1715
1605 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1664
1716 CTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1763
1665 CATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1724
1764 CAGAGTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1823
1725 GAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1784
1824 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1883
1785 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1844
1884 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1943
1845 GTC-----GCTTCGCTGCGCGGCTGCAAGCTGGAAGGCTGCGCGCGGAGG 1895
1944 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2003
1896 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1955
2004 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2063
1956 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2015
2064 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2123
2016 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2075
2124 GAGCAGTCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2183
2076 CTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2135
2184 CAAGATGCTCGGCTA 2198
2136 CAAGATGCTCGGCTA 2150

RESULT 6
US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 636837
; GENERAL INFORMATION:
; APPLICANT: Sarsisani, Sina F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bio-production of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A

CURRENT FILING DATE: 2000-07-27
 PRIOR APPLICATION NUMBER: 68/147,719
 PRIOR FILING DATE: 1999-08-06
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: Microsoft Office 97
 SEQ ID NO: 7
 LENGTH: 2151
 TYPE: DNA
 ORGANISM: Rhodococcus glutinis
 US-09-627-216A-7

Query Match 49.4%; Score 1195; DB 4; Length 2151;
 Best Local Similarity 73.9%; Pred. No. 7, 8e-223;
 Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

QY 99 CGCCGCTCCGACCAAGTGGCTGGGGGCCCCACTTGGGCTTCGCGCCGACGCCGCT 158
 DB 45 CGTGGCATCGCAAAACAGGCTGTCAATGGCGCTTCGACCACTCGCAGTCGAGGCTC 104
 QY 159 CGATGGCAAGCGCGCACCAAGTCGAGCTCGAGATTCGAGAGAGAGGCTCTCAGAGCC 218
 DB 105 GCACTGCCCAAGACCCAGGTACAGGTCGACATTCGAGAGAGATGCTTCGCGGCC 164
 QY 219 CACCGACAGCTGTCGAGCTCAGCGGTACAGCTCAGCTCGGTGAGCTTGTGCGGC 278
 DB 165 GACCGACTGACGCTGAACTCGACGGCTACTGCTCAACCTCGAGAGCTGTCTGCGC 224
 QY 279 CGCCGCAAGGGGCGGAGGCTCGGCTCCAGAACGACGACGAGATTCGCGGACGCTCGA 338
 DB 225 CGGAGAGAGGAGGAGGCTGTCCGGTCAAGGACGACGACGATCGCTCAAGATTTGA 284
 QY 339 CAAAGCGTCACTTCTTCAGAGCCAGCTTCAGAACTCGGTCTACGAGTCAACAGCG 398
 DB 285 CAATCGGTGAGTTCCTTCGCTCGCACTTCGATGAGGCTTACGGCTCAGGACTGG 344
 QY 399 TTTCGGTGGCTGGCGGACAGAGACTGAGAGATGCACTCAGCTCCAGAGGCGCTCAT 458
 DB 345 ATTGGCGGATCCGAGACACCGGACGAGAGACCCATTCGCTCCAGAGGCTCTCT 404
 QY 459 CGAGCACAGCTCTCGGCGTGAAGCGCAGCTCGGCTCGCTTCAGCGCTCGAGAGCGG 518
 DB 405 CGAGCACAGCTCTCGGCGTGAAGCGCAGCTCGGCTCGCTTCAGCGCTCGAGAGCGG 464
 QY 519 CTTGAGAACAGAGCTTCGCTGAGTGTGTGCGGCGCAATGCTATCCGCTCAATC 578
 DB 465 TCTCGAGAACTGCTTCCCTCGAGTGTGTGCGGCGCAATGCTATCCGCTCAATC 524
 QY 579 GCTCAAGCGTGGGCACTCGGCGGTGCGGCTGCTGCTTGAAGCGCTCAACATTTCT 638
 DB 525 CTTGACCGGCGGCACTCGGCGGTGCGGCTGCTGCTTGAAGCGCTCAACATTTCT 584
 QY 639 GAAACACGACATCAGCCCATCTGTCGCCCTCGGCGCTTCATCTCGGCTCGGAGCGT 698
 DB 585 CAACACGACATCAGCCCATCTGTCGCCCTCGGCGCTTCATCTCGGCTCGGAGCGT 644
 QY 699 CAGCCGCTTCGTAATCGCGCGGCGCATTCAGCGGTCAACCGGCTCAAGCTTACGT 758
 DB 645 GTCTCTCTCTCTCAATTGACGCGCATTCAGCGGTCAACCGGCTCAAGCTTACGT 704
 QY 759 TTTGACAGAGGAAACGAGAGATCATGTTGCGCGGAGGCGGCTCATCGCTTTTGGTCT 818
 DB 705 CGTTCACAGAGGAAACGAGAGATCATGTTGCGCGGAGGCGGCTCATCGCTTTTGGTCT 764
 QY 819 CGAGGCAATGCTCTCTCGGCGGAGAGAGGCTTGGCTTGTGTCACGAAACGCGCTCTC 878
 DB 765 CGAGGCGCTCTCTCTCGGCGGAGAGAGGCTTGGCTTGTGTCACGAAACGCGCTCTC 824
 QY 879 CGCCTCATGCGGACCTTCAGTGTGACGACTCGACATGCTCGCTCTCTCGGAGGCG 938
 DB 825 AGCATCATGCGGACCTTCAGTGTGACGACTCGACATGCTCTCTCTCTCGGAGGCG 884
 QY 939 CTTACGCGCTCTCAGGTCAGGCGCATGTGCGGCGAGAGGCTCTGTCGCGCTTCAT 998

DB 885 GCTACAGGACATGACGATCGAAGCGATGTCGCGCACCGCGCTGTTCACCCCTTCT 944
 QY 999 CCAGCAAGCTGTGCGCGCGCACCCCGGCGAGGTGCGGCGCAACATCCGACGCT 1058
 DB 945 TCACGACGTCAGCGCGCTTCACCCGACGACATGAAAGTGGGGGAAACATCCGACGCT 1004
 QY 1059 CTTTTCGCGCTGTGCTTGGCTTGAACAGAGAGAGTCAAGGTCAAGAGACGACGA 1118
 DB 1005 CTTGAGAGGAGCCGCTTGTGCTACCATGAGAGAGAGTCAAGTCAAGAGACGACGA 1064
 QY 1119 GGGCAATCTTTCGCGGAGACCGCTACCGCTCCGACGTCGCTCAATGCTTCGCGCT 1178
 DB 1065 GGGCAATCTTTCGCGGAGACCGCTACCGCTCCGACGTCGCTCAATGCTTCGCGCT 1124
 QY 1179 CGTGGAGGACATGATGACAGGCTTACGCTCTCTCGTCAAGACAAACA---GACGAC 1235
 DB 1125 CGTCAAGGACCTCATTCAGCGCTCAAGCGCGCTCAGCATGAGGCGCGGACGTCAGAC 1184
 QY 1236 CGACAAACCGCTCTCTCGACGTCGAGAACAGACAGACCGGACGCGCAATTCCAGCT 1295
 DB 1185 CGACAAACCGCTCTCATTCAGCGCTCAAGCGCGCTCAGCATGAGGCGCGGACGTCAGAC 1244
 QY 1296 GTGCGCTGTCTCATTTGATGAGAGACGAGGCTGCACTGCGCTCATTCGCGCAAGCT 1355
 DB 1245 TGGCGCTGTGCGCAACACATGAGAGAGCTCGCTTGGGCTGCGGCAATTCGCGCAAGCT 1304
 QY 1356 CAATTCACGAGTGCACCGAGTGTCTCAACGCTGCGCATGACCGCGCTGCTTCTGT 1415
 DB 1305 CAATTCACGAGTGCACCGAGTGTCTCAACGCTGCGCATGACCGCGCTGCTTCTGT 1364
 QY 1416 CTTGCTGCGGAGGACCGGTGCTCACTATCAAGGAGGCTTGAACATTCACATTCG 1475
 DB 1365 CTTGCTGCGGAGGACCGGTGCTCACTATCAAGGAGGCTTGAACATTCACATTCG 1424
 QY 1476 TGTTCAGCTTCGAGTTCGCGCACTTTCGCAACCGGCTCACTACCTTGTCTCGAGCGCG 1535
 DB 1425 GCGGTACACCTCGAGTTCGCGCACTTTCGCAACCGGCTCACTACCTTGTCTCGAGCGCG 1484
 QY 1536 AGAGATGGGACACGAGCGGCTTCACTTCGCTGCTTCATTCGCGCGCGGCACTGCGCA 1595
 DB 1485 TGAATGGGACACGAGCGGCTTCACTTCGCTGCTTCATTCGCGCGCGGCACTGCGCA 1544
 QY 1596 GCGCAAGAGCTCTTCT 1655
 DB 1545 GTCACAGAGCTCTTCT 1604
 QY 1656 GAGCTTCGCGCGAGTGAAGCTTCACTTCAAGAGAGTTCGAGCGGCTTCTCTCTCTCT 1715
 DB 1605 GAGCTTCGCGCGAGTGAAGCTTCACTTCAAGAGAGTTCGAGCGGCTTCTCTCTCTCT 1664
 QY 1716 CTTGCAAGAGCT 1763
 DB 1665 CATGACACGACATTTGGCTTCGCGCATGACCGCTTGAACCTGCGGACGAGCTGTGGA 1724
 QY 1764 CAGGTCAGAGAGCGCTTCAACAGCGTTCAGACGACGACGACGACGACGACGACGAC 1823
 DB 1725 GAGGTGACAGAGCGCTTCAACAGCGTTCAGACGACGACGACGACGACGACGACGAC 1784
 QY 1824 GCGGTGACAGAGCGCTTCAACAGCGTTCAGACGACGACGACGACGACGACGACGACGAC 1883
 DB 1785 GCGGTGACAGAGCGCTTCAACAGCGTTCAGACGACGACGACGACGACGACGACGACGAC 1844
 QY 1884 GCGCTTCGCAACGTCACCTTACTGCTCAACGCTGAGAGGTTGCTGCGCGGAGAA 1943
 DB 1845 GTC-----GCTCTGCTGCGCGGCTTCAACGCTGAGAGGTTGCTGCGCGGAGAA 1895
 QY 1944 GGCATCTGCTTCAAGCGCGGAGGTGAGAACCGCTTTCGAGAGACCGCTTTCGAGAGC 2003
 DB 1896 GGCATCTGCTTCAAGCGCGGAGGTGAGAACCGCTTTCGAGAGACCGCTTTCGAGAGC 1955
 QY 2004 GCGGCGGACGACATCTCTGCGCGGACGCGGCTCTGACTGCTGCTGCTGCTGCTGCTGCT 2063
 DB 1956 GCGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015

QY 1656 CGACCTCCGCGAGATGAGAGTGCCTTCAAGAAAGAGTTGACCCGCTTCTCCGACTCT 1715
DB 1605 CGACTTCGCGCGATGAGTTGAGTTCAAGAAAGAGTTCCGCTCCGCTCCGCT 1664
QY 1716 CCTCCAGCAGACCTTCGCGACTGCGCTCCGCTCA-----ACGCACTTGGCGCT 1763
DB 1665 CATGACACAGCACTTTGGCTCCGCGCATGACCGGCTGAAACCTGCGCAGACGCTTCGCA 1724
QY 1764 CGAGCTCAAGAAAGCGCTCAACAAGGCTTCGAGAGAGAGAGCTTACGACTTCGAGCC 1823
DB 1725 GAAGGTGAACAAACGCTCGCCAGAGGCTCGAGAGAGACCAACTGCTGACGCTGCTCC 1784
QY 1824 GCGCTGAGCAGACGCTTCTGCTACGCGACCGGACCGTCCGAGCTCTCTGCTCTC 1883
DB 1785 GCGCTGAGCAGAGCGCTTCTCTTCCGCGCGGACCGCTGCTGAGGCTCTCTGCTGAC 1844
QY 1884 GCGCTTGCACAGCTACCTTACCTGCTGACGCTGAGAGGTTGCTCGCGCGAGAA 1943
DB 1845 GTC-----GCTCTCGCTCGCGCGCTCAACGCTGAGAGGTCGCGCGCGAGTC 1895
QY 1944 GGCATCTCGCTCAAGCGGAGAGTGGGAACCGCTTCGAGAGACCGCTTTCGAGGC 2003
DB 1896 GGCATCTCGCTCAAGCGGAGAGTGGGAACCGCTTTCGAGAGACCGCTTTCGAGGC 1955
QY 2004 GCGCGGAGCAGATACCTCTGCGCGGAGCGGCTCTGACTGCTTCTGCGCGAGAA 2063
DB 1956 GCGCGGCTCTGCTGACTCTGCGCGGAGCTCAAGCTCTGCTGCTGCTGCTGCTGAG 2015
QY 2064 GCTCGCGCTGAGAGGCGCGCGGAGAGTGTGCTGCTGAGAGAGAGATCGG 2123
DB 2016 GCTGCGCTGAGAGGCGCGCGGAGAGTGTGCTGCTGAGAGAGAGATCGG 2075
QY 2124 GAGCAAGCTCTGCGCATCTACAGAGGCGATCAAGAGAGCGGCTCAACAGCGCTCG 2183
DB 2076 CTCGAAGCTCTCAAGATCTAGAGGCGATCAAGTCCGCGGAGATCAACAGCTCTCT 2135
QY 2184 CAAGATGCTCGCTA 2198
DB 2136 CAAGATGCTCGCTA 2150

RESULT 8

US-09-765-873A-31
Sequence 31, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BCI009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 31
LENGTH: 2061
TYPE: DNA
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: CDS1
LOCATION: (1)..(2061)
US-09-765-873A-31

Query Match 49.1%; Score 1188.8; DB 4; Length 2061;
Best Local Similarity 75.2%; Pred. No. 1.2e-221;
Matches 1529; Conservative 1; Mismatches 478; Indels 24; Gaps 3;

QY 182 CGCAGCTCGAGATGTGACAGAGCTCTCAGCAGCCCAACGAGAGCTGTGAGCTCA 241
|||||
DB 118 AGAACAAGACTTCGACCAAGCGGCAATTTCAAGGCTGCGCTGTGCTCAACACATGG 1177

DB 38 CCGAGGTGACATCTGTGAGAAATGCTCGCGCGGCGGAGGAGTCAAGCTGCAACTCG 97
QY 242 GCGGAGTACAGCTTCAAGCTTCCGTGAGCTTGTGCGGCGGCGGAGGAGGAGGAGTCC 301
DB 98 ACGGCTACCTCGGTCAAGCTTCGAGAGAGTGTGCTGCGCGGAGAGGAGGAGGCTGTCC 157
QY 302 GCGTCAAGAGAGAGAGAGAGATTCGCGGAGCGCTGCAAGAGAGGCTGACTTCTTAAG 361
DB 158 GCGTCAAGAGAGAGAGAGATTCGCTCAAGATTAAGTCAATGCTGAGTCTTTCGCT 217
QY 362 CCGAGCTTCAAGAGAGAGAGTCAAGAGTCAAGAGGTTTGGAGTTCGCGGAGAGCA 421
DB 218 CGCAACTCTCATAGAGGTCTACGCGCTCAAGAGTCAAGATTTGCGGATTCGAGAACCC 277
QY 422 GAGCTGAGAGATGAGTCAAGCTTCAAGAGGCGCTCATGAGACCAAGCTTCGCGGAG 481
DB 278 GCACCGAGAGAGAGAGTCTGCTCCAGAGAGGCTCTCCAGAGCAAGCTTCGCGGAGTTC 337
QY 482 CCGCAGAGTCCGCTGCTTCAAGGCTGAGAGGCGGCTCGAGAACAGCTTCGCTG 541
DB 338 TCCCTTGTGCTGACTGCTTCCGCTTCGAGCGGCTTCGAGAACTGCTTCCCTG 397
QY 542 AGGTGTCGCGGCGGAGAGTCAATCCGCTCAACTGCTCAAGCTGAGCACTCGAGCG 601
DB 398 AGGTGTCGCGGCGGAGAGTCAATCCGCTCAAGCTTGAACCGGCTCACTGAGCG 457
QY 602 TCCGCTGCTGCTTGAAGGCTTCAAGCTTGAAGCAAGGCTCAAGGCTCAAGGCTCA 661
DB 458 TCCGCTGCTGCTTGAAGGCTTCAAGCTTGAAGCAAGGCTCAAGGCTCAAGGCTCA 517
QY 662 TCCGCTGCTGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 721
DB 518 TCCGCTGCTGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 577
QY 722 GCGCTCAATCCGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 781
DB 578 GCGCTCAATCCGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCAAGGCTTCA 637
QY 782 TCATGTTGCGGAGAGGAGATCTGCTTTCGCTTTCGAGAGAGTTCGCTTTCGAGAG 841
DB 638 TCTGTAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTC 697
QY 842 AGAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAG 901
DB 698 AGAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAG 757
QY 902 TGCAAGCTGCGACATGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGG 961
DB 758 TGCAAGCTGCGACATGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGG 817
QY 962 CATGCTGCGCAGAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGG 1021
DB 818 CATGCTGCGCAGAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGG 877
QY 1022 CCGGCTGAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTC 1081
DB 878 CCGGCTGAGAGGCTTTCGAGAGGAGATGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTC 937
QY 1082 TTGAGCAGAGAGAGAGATCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 1141
DB 938 TTGAGCAGAGAGAGAGATCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 997
QY 1142 ACCGCTTCCGAGAGGCTTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 1201
DB 998 ACCGCTTCCGAGAGGCTTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 1057
QY 1202 ACTGAGCTCTGCGAGAG---AACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1258
DB 1058 ACCGCTTCCGAGAGGCTTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 1117
QY 1259 AGAACAAG 1318
DB 1118 AGAACAAG 1177

QY	867	AAAGGCGCGTTCGAGCCCTGATGAGCGAGACCCCTCACTGCAAGACTGCGCAATAGTCTTCGCT	928
Db	1464	CAGGCGCGTTCGAGCCCTGATGAGCGAGCGCGCCCTCAACCGAGCGACAGCTCTTCGCT	1522
QY	927	CCTTCGCGAGCGCTTGACGGCTCTCAACGGTGAAGGCATATGTGCGGCACAGAGCGCTGCT	986
Db	1524	CCTCGACAGAGCGCCTCACTGCTCTTACTGTGAGGCGCATATGTGCGACAGCGCGCTGCT	1582
QY	987	CGGCGCGTTCATCAGAGAGTCTGCGCGCCCGACACCCCGGCGAGGTGAGGTGCGCGCA	1046
Db	1584	CACCCATTCTCTCAAGAGTCAACGGGCTTCAACCGAACCGAGATGAGGTGGCGGCA	1642
QY	1047	CATCGCAGCGTCTCTTCGCGGCTCGTGTTCGCTTACCGTTGAGCAGAGGAGAGGTCAAGT	1106
Db	1644	CATCCGGACTCTTCTCGAGGGGACGAAAGTAAAGCCCGTCCACACAGAGACTGAAGTCAAGT	1702
QY	1107	CAGGACGAGAGGGGATTTCTTGCCAGGACCGCTAACCCGCTCGGACGTCGCTCAAGTT	1166
Db	1704	CAGGACGAGAGGGGATCTCAGGAGAGACCGGTAAACCGCTCGCTGCTCGCCAGATG	1763
QY	1167	CGTCGCGCGCTCGTGAAGACATGATGACCGCTTACCTGCTCTGCTGCGAG--AA	1223
Db	1764	GCTCGGTCCCTTGTGAGGAGATATTTACGCTACGCTGTCTCTGCTGAGGCTGCG	1823
QY	1224	CAACACGACGACCGACCAACCCGCTCTCTGACGCTGAGAACAGACGACCGCGACCGCG	1283
Db	1824	TCAGTGACACACCGACCAACCCGCTATGACCTCGAGAACAAAGATGACCCACATAGCGG	1883
QY	1284	CAACTTCGAGGGCGTGGCTGTGTGATTTGATGAGAGACAGAGCTGCACTGCGCT	1343
Db	1884	AACCTTCATGAGGAGACGCGTGTGAACAAGATGAGAAAGCTGCGCTGCGCGT	1943
QY	1344	CATCGGCAAGCTCAACTTCACGCAATGACAGAGTGTCTCAAGCTGACATGAACCGCG	1403
Db	1944	GATGGCAAGGTCACTTTTACTACGCTACCGAAGTGTCAACCGCGGCAATGAACCGCG	2003
QY	1404	CCTGCTTCGTCCTTCGCTGCTGCGCGAGAACCCGTGCTCAACTATACCGGCAAGGCTTGA	1463
Db	2004	CCTTCGCTGCTGCTGCTGCTGCGAGAACCTTCCTCTCTTATACATGCAAGGCTTCGA	2063
QY	1464	CATTCAATCGCTGCTTACGCTTCGAGACTGAGCCACTTGTCCGAACCGGCTCACTACTT	1523
Db	2064	CATTGCTGGGCGCGCTACACTTCGAGACTGCTCACTTGTCCGAACCGGCTTGAACCA	2123
QY	1524	CGTCAAGCCCGACAGATGGGCAACAGAGCCGTAACCTGCTGCTTCATATTCGCGCG	1583
Db	2124	CGTCAAGCCCGACAGATGGGCAACAGAGCCATCAACCTGCTGCTCAATTCGCGCGG	2183
QY	1584	CGGCACTCCGAGGCGCAAGCAAGCTCTTCTCTCTGCTGCGCCGTGCAAGCTGATGAC	1643
Db	2184	CGGCACTCCGAGGCGCAAGCAAGCTCTTCTCTCTGCGCCGTGCAAGCTGATGAC	2243
QY	1644	GCTCAAGGCGCTGCACTTCGCGCGAGATGAGACTGCACTTCAAGAGCAAGTTGACACCGCT	1703
Db	2244	CCTCAAGGCGCTGCACTTCGCGCGAGATGAGATTGACACCAACCAAGCGTTGAGCGGAT	2263
QY	1704	TCTCCGAGCTTCCCTCCAGAGGACACTGCGACTGCGCTGAGAGTCAAGCACTTGCGCT	1763
Db	2304	GGTACTAGAGCTGTGAAGAGCACTTTGGC--GCGCTGCGAGCGCTCAAGTTCAGAGA	2360
QY	1764	CGAGGTCAAGAGGCGCTCAACAGAGCTTCGACAGACAGCAAGTACAGCTTCAGGCT	1823
Db	2361	CAGGTCCGCAAGTGCATTAACAAGCGGTGACAGCAACAACCTGTAAGACTTCAGACA	2420
QY	1824	GCGGTGCGAGAGCGCTTCTGTAACGAGACCGGACCGGTGTGAGGCTCTCTTCGCTTC	1883
Db	2421	GCGGTGCGAGAGCGCTTCTGTAACGAGACCGGAGCGGTGTGAGAGCGCTTCGCGGCA	2480
QY	1884	GCCCTTCGCAACGTCACCTTACTGCGCTCAACGCGGTGAAGAGTTGCTCGGCGAGAA	1943
Db	2481	G-----GAGGTTCGCTCGGAGACTTCMAAGCTGGAAGTGGCTCGGCGAGAA	2531
QY	1944	GGCATTCGCTACGCGGCAAGTGGCGCAACCGCTTTCGAGACAGCCGCTTTCGAGGCT	2003

Db	2532	GGCTATTCGGGCTCAAGCGGCTCCGTCGCGGACCTCGTTCTCGGGCGAGTCCGTCGTCGTCGTC	2591
QY	2004	GGCGGCGGACGAGATACCTACCTCCGCGCGACGCGGCTCTGTACTCGTTCCGCGGAGGA	2063
Db	2592	GCGCGGCTCAAGTACTCTCTCCCGCGGACGCGGCTCTGTATTCTTCGTCGCGGAGGA	2651
QY	2064	GCTCGGCGGACAGGCGCGCGCGGCGAGGTATTGTCCGTCGTGACGAGGAGACGATCG	2123
Db	2652	GCTCGGCGTCAAGGCGCGCGCGCGGAGTGTCTACTCGGCAAGGAGGAGTCAACGATCG	2711
QY	2124	GAGCAAGCTTCGCGCATCTACGAGGCGCATCAGAGCGCGCATCAACCGTCCTCGT	2183
Db	2712	CACCAACTCAGCGGCATCTACGAGGCATCAGAGCGGTTCATCGCCCCCGTCTCGT	2771
QY	2184	CAGAGTCTTCGCGTA	2198
Db	2772	CAGGATGATGCGATA	2786

```

RESULT 10
US-09-624-693A-14
/ Sequence 14, Application US/09624693A
/ Patent No. 655468
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ APPLICANT: Koostra, Anna
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ TITLE OF INVENTION: Using Same
/ FILE REFERENCE: 29479/500NSC
/ CURRENT APPLICATION NUMBER: US/09/624,693A
/ CURRENT FILING DATE: 2000-07-24
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 14
/ LENGTH: 2311
/ TYPE: DNA
/ ORGANISM: Amanita muscaria
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (18)..(2237)
US-09-624-693A-14

```

Query Match	12.5%	Score 302.8	DB 4	Length 2311
Best Local Similarity	54.2%	Pred. No. 4.5e-50		
Matches 719	Conservative	1	Mismatches 573	Indels 33
				Gaps 4
QY	379 GTCTACGAGTACACCGAGGCTTTTCGAGTCTCGCGGCACACGAGACTGAGATGCGACATC	438		
Db	372 GATATAGAGCTGTCAACTGGTTTCGGTGGCAGTGTCTATATACAGGAGGACAAACCGATG	431		
QY	439 AGCCTCAGAAAGGCGCTATCAGACACCACTCTGCGCGCTGACGCGCAGCTTCGCGCTG	498		
Db	432 TTGTTGGGTTTGGCCCTTTTGCAACACCAACATGTAGGATTACTGCCCACTCGACTGAG	491		
QY	499 TCCTTCAGGGTGGACGCGGCGCTG-----AGAACCGCTTTCCTCGAGTGTGTC	549		
Db	492 CCTTTGAGAGTCTTCACTCTTCAAGATGAAATTAACAAAGCATGCGAAGGCGTGAGT	551		
QY	550 CGCGGCGCCATGTCTATCCGCGTCACTCGCTCAGCGGTGGCCACTCGGCGCTCGCGCTC	609		
Db	552 CCGGGGGCCATTTTGATCCGATTAATTCGCTAATTCGTGGCCACTCTTGGAATTCAGATG	611		
QY	610 GTTCGCTCTTAGAGGCGCTTACCACCTTTTGAACCAACGCGATCAAGCCCATGTCGCCCTC	669		
Db	612 GAGTTGATCGAAAGATGAGAGAACTCTCGCGGCAATGTATACCTGTCTTCCCTG	671		
QY	670 CGCGGCTCATCTCGGGGTGGGGGAGACCTCAGCGCGCTCTGTATCATCGCCGCGCCATC	729		
Db	672 AAGGAGAGATCTCTCATATCCGAGATGTGTCTCCCTTATCTCATATTCGACGGCAGATT	731		
QY	730 ACCGCTACCCCGACGTCAAGTTCA-----GTTTGCACGAGGGAACACAGAGATC	783		

Db ATTGGCAACCATCAATCAAGTATATACAGGTCCATCAAAAGCCGGAATTCGCCAAATT 791
732 ATTTGGCGCGAGGCGCATCTCGCTCTTTGTGTGAGGCACTGCTCTCGGCCGAAG 843
784 ATGTTGGCGCGAGGCGCATCTCGCTCTTTGTGTGAGGCACTGCTCTCGGCCGAAG 843
792 GGATCTCGAAGATGCTTGGCTCTGCAATATATCAAACTTTCCACATGAGATGAAA 851
844 GAGGCTCTCGGCTCTGCTCAACGGAAGCGCGCTCGGCTCGATGGGAGACCTCACTG 903
852 GAACCTCTGGATTTTGAATGGAGCGCAATTCCTGCAATCTGTGGACGCTTAAAGCTTA 911
904 CACGACTCGCATGCTCTCGCTCTCTCGAGGCTTGAAGGCTCTCAAGGTGAGGCTC 963
912 AACGAGCTATCCATCTGCTTGTGTGTGCTCAAGTGTGACGCGCTAAGGAGACGAGGA 971
964 ATGTCGCGCCAGCAGGCTCTGTGTGCGCGCTCTCAACGAGCTGTGCGCGCGCGACCC 1023
972 TTGATAGGCACTCGGCTCTCTCAATGCAACGCTTCAATATCCACCGCAACACATCC 1031
1024 GCGCAGTCTGAGGTCTGCGCGCAACATCCGACGCTCTCTCGGCTGTGCTTGGCTT 1083
1032 GGTCAAGTGAATGTGCTGAGAACATTTGGATTTGCTGATGGAGTAAATTGGCTCAG 1091
1084 GAGCAGAGGAGAGGTCAAGGTCAAGAGCAGAGGCAATTTTGGCCAGGACCGCTAC 1143
1092 TTGAGAGAGCAGAGTTCGCTTGAAGACGATTAATACCCCTTGGCAGGACCGTTAT 1151
1144 CGGCTCGCGACGCTCGCTCAATCTCTGCGCGCGCTCTGAGAGCATGATGACGCTAC 1203
1152 CCATCTCCAACTTCCCTCAATTCCTTGTGGCTTCAAGTGAAGCATTAATCTCGCTTC 1211
1204 TCGACTCTCTGCTGAG-----AACAAACGAGCAGACCGAACCCGCTCTGAGCTC 1257
1212 CAGACTGTAAACGAGAGTGTAAATTAATTAACCACTACTGCAATCCAGTAAATGATG 1271
1258 GAGAAACAAGCAGACCGCGCAGCGCGCACTTCAGAGGCTGTGCTGTGATTTGGATG 1317
1272 GAGACTGGGATCTCACCGAGTGGCAATTTCCAGAGATGCTGTAACTAATGCAATG 1331
1318 GAGAAAGCAGGCTGCACTGCGCTCTCACTCGGCAAGCTCAATTCAGCAGATGACCGAG 1377
1332 GAGAAAGCAGGCTGCTTGTAACTAATTAATTAATTTCCAGAGACTGAA 1391
1378 TTGCTCAACGCTGCTCAATGACCGCGCTGCTGTGTGCTGTGCGAGAGACCGCTCG 1437
1392 TTAGTCAATCTGCGATGAACCGCGCTGTGCGCTTCAAGTGTGCAAGATCCATCT 1451
1438 CTCAACTATCAAGGAGGAGGCTTGAACATTCACATCGCTTCAAGCTGAGCTCGAG 1497
1452 CTCAACTATCAAGGAGGAGGCTTGAACATTCAGCACTGCGCTTCAAGCTGAGCTCGAG 1504
1498 CACCTTGGCCAAACCGGCTCACTACCTTGTGCAAGCGGAGATGGGCAACAGGCGCTC 1557
1505 -----AGGACCTCTGCGCGCGCACTCACTTCACTGAGGAGAAATGCAACCAAGCTGTT 1559
1558 AACTGCTGCTCTCACTCTCGCGCGCGCGCACTGCGGAGGCAACAGCTCTTCTCTC 1617
1560 AACTCTCTGCTGCTTCACTCTGCTGCGGCTCACTCACTGCTTGAAGTGAACATCT 1619
1618 CTTCCTGCGCTGCACTGCTTCACTGCAAGCTGCAAGCGCTGCACTCGCGCATGAGCTC 1677
1620 CTGATGCGCTCTTACTGTTAATTTATGAGCAGAGCTCTGAGCTCTCGAGCTTCAAGGCG 1679
QY 1678 GACTTC 1683
Db 1680 GAGTTC 1685

APPLICANT: Bloksberg, Leonard N.
APPLICANT: Hayakkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
TITLE OF INVENTION: Modification of Plant Lipid Content
FILE REFERENCE: 11000.10034U
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 245
LENGTH: 1455
TYPE: DNA
ORGANISM: Eucalyptus grandis
us-09-615-192a-245
Query Match 8.6%; Score 207.2; DB 4; Length 1455;
Best Local Similarity 52.4%; Pred. No. 1,4e-31;
Matches 623; Conservative 0; Mismatches 528; Indels 39; Gaps 6;
QY 184 CAGCTCGAGATCTGTCGAGAGCTCTCAGCGACCCCAACGACGATGCTGAGCTCAG 243
Db 299 CACCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 358
QY 244 GGGTACAGCTCAGCTCGCTGAGAGTGTGTGCGCGCGCGCGCGCGAGAGGAGGAGGAGGAG 303
Db 359 GGGAGTCCCTCAGAGATAGCCAGAGTGGCGGCGGTGCGAGTGAAGAG--GGGATAGGG 415
QY 304 GTCCAGAAAGCAGAGAGATCGCGGACGCTGTGACAAAGGCTGTCACTTCCCAAGGCC 363
Db 416 GTCCAGCTCTGAGAGGCG 475
QY 364 CAGCTTCAGAA-----CTGGGTACGAGATCCACACGAGGCTTGTGAGCTCGGCGGAC 417
Db 476 AGCATGAACAGGAAATCGAGAGTACAGGAGTACAGGAGTACAGGAGTACAGGAGTAC 535
QY 418 ACGAGACTGAGAGTACAGTACAGCTCCAGAGGCTGTATCGAGACCAAGCTCTGCGGC 477
Db 536 CGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 594
QY 478 GTGAGCGGAGCTCGCTGTCTCTTCAAGCTGTGAGAGCGCGCGCTCGAGAAACAGCTTCCG 537
Db 595 -----GATCTTGGCAACGCGACGAGGCTGTGCAACACCTTCTCT 634
QY 538 CTGAGAGTCTCG 597
Db 635 CAATCTCTCAGCGAGCG 694
QY 598 GCGGCTCGCTGTGCTGTGAGGCGCTCAAGCTTCTTGAACCAACCGATACCGCC 657
Db 695 GGCATCTCGTTTGAATCTCGAGGCGCATCAAGTCTCTCAACCAACATCACCCCG 754
QY 658 ATCTGCTCTCG 717
Db 755 TGCCTGCGCTTGAAGGAGCAATCACTGCTCAAGGCGAGCTTGTGCTCTCTCTCAAT 814
QY 718 GCGGCGCGCATCACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
Db 815 GCGGCGCTCTGAGCG 871
QY 778 AAGATCATGTTTGGCGCGGAGGAGCATCTCGCTCTTGTGTCTCGAGGAGGAGTGTCTCGGC 837
Db 872 CTGAGCGCTGTGAGGAGCTTTCGCTGTGCGGAGTGAACAGGAGCTTCTTGAAGTGAAG 931
QY 838 CCGAAGAGAGGCTCTGAGTGTGTGCAACGAGAGCGCGCTCTCGCTCGATGAGCGACCTTC 897
Db 932 CCAAGAGAGGCTGTGCGCTGTGAGACGAGCAGCGAGTGTGAGTGTGAGCTTCAATC 991

RESULT 11
US-09-615-192a-245
; Sequence 245, Application US/09615192a
; Patent No. 6410718
; GENERAL INFORMATION:

TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
J5-09-252-991A-16475

Query Match 5.2%; Score 126.8; DB 4; Length 1548;
Best Local Similarity 47.1%; Pred. No. 5.1e-16;
Matches 571; Conservative 0; Mismatches 592; Indels 49; Gaps 4;

2Y 516 CGGCTCGAGAACAGCTTCGCTCCGCTGAGCTGTCGCGCGCGCATGTCATCCGGCTCA 575
DB |||||
351 CGGCTCGAGAACAGCTTCGCTCCGCTGAGCTGTCGCGCGCGCATGTCATCCGGCTCA 410
2Y CTGCTCAGCGCTGAGCACTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 635
DB |||||
411 CAGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 470
2Y 636 CTTGAACCCAGCATACAGCCCATGCTCCCTCCGCGCTTCATCTCGCGCTGCGGCA 695
DB |||||
471 GATCAACCGCCAGAGTTCATCCGACATCCCGTGAAGGCTCGGTGGTGGTCCGCGCA 530
2Y 696 CCGTACGCGCGCTCTGATCATCGCGCGCGCATCAGCGCTCAGCCGACGCTCAAGGTTCA 755
DB |||||
531 CCGCGCGCGCTGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 590
2Y 756 CGTTTTCAGCAGGGAACCGAAGATCATGTTTGGCGCGCGAGGCTCTGCTCTTTGG 815
DB |||||
591 CGGTG-----AGTGGCTGCCGCGCGCGCGAGCGCTGCGGTGGCGG 632
2Y 816 TCTCGAGGCACTGCTGCTGCGCGCGCGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
DB |||||
633 GCTGAGCGCGCTGACCTGCGCGCGAGAGAGGCTTGGCTGCTGCTGCTGCTGCTGCTGCT 692
2Y 876 CTGCGCTCGATGAGGCACTCTGATGCTGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTG 935
DB |||||
693 GTCCACCGCTACGCGCTGCGCGCGGTTTTCAGAGCGGAGACTGTTCCCGCGCGAC 752
2Y 936 GGCCTTACGCGCTTACGCTGAGAGGCTGCTGCGCGAGAGGCTGCTGCTGCTGCTGCT 995
DB |||||
753 GGTGCGCGCGCGCTCAGCGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 812
2Y 996 CATTCAGACGCTGCTGCGCGCGCGACCGCGCGAGTGCAGGCTGCGCGCAATCCGCA 1055
DB |||||
813 CATTCATGCGCGCGCGCGCGAGCTGCGCGAGATGACATGCGCGCGCTGCTGCTGCTGCT 872
2Y 1056 GCTGCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1115
DB |||||
873 GCTCAACCGCAAGACGAGGTGGCGCGCTCCCATGAGAGTGCACAAGT----- 923
2Y 1116 CGAGGCACTTTCGCGCGAGACGCTACCGCTCCGCACTGCTGCTGCTGCTGCTGCTG 1175
DB |||||
924 -----CGAGACCGCTTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 968
2Y 1176 GCTGCGAGGACATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1235
DB |||||
965 CTGCTGACCGCAAGTGGCGCGCGCGCGCGAGTGTGAGATGCAACCGAGGTGTC 1028
2Y 1236 CGACAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1295
DB |||||
1029 CGACAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1088
2Y 1296 GTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
DB |||||
1089 CGAAGCGGCGAGAGCGCGCGCGCACTGCGCTGCGCTGCGAGATCGGTTGCG- 1147
2Y 1356 CAATTTCAGCAGTGCAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
DB |||||
1148 -----TGTGGAAGCGCGCATCTGCTGATGATGAGACATGACATGTCAGAGTTGCG 1202
2Y 1416 CCGTGGTGGAGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1475
DB |||||
1203 GTTCCGTGGGCGAAGCGCGGCGGCTCACTCGCGCTTCAATGATCGCGCGAGTCA 1262
2Y 1476 TGGTTACGCTTGGAGCTGCGCGACCTTGCACCGCGTCACTTCTGCTGCGCGCG 1535
DB |||||

1263 CGGCTCGAGCAACAGAGCGCTGCGCGCATCCGCGAGGCTGAGAGCTGCGCGAC 1322
DB |||||
1336 AGAATGAGCAACAGCGCGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1595
DB |||||
1323 CTGCGCAACCGAGAGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1382
2Y 1596 GAGCAACGAGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1654
DB |||||
1383 GATGCGCGAGAGAGTCCCGCGGATTCGCGCGTGGAGTGGCTGGGCGCTGCGAGGCGC 1442
2Y 1655 TCGACCTCGCGCGAGTGAAGCTGACCTTCAAGAGAGCTTCAACCGCTTCCCGACTC 1714
DB |||||
1443 TGAAGCTTCCGAGGCGCTTGAAGAGTTCGCGAGCTGAGAGAGCTGCGCGCTGCTGCT 1502
2Y 1715 TCTTCAGAGAGC 1726
DB |||||
1503 GCGCAAGGTC 1514
DB |||||

RESULT 14
US-09-252-991A-16126/c
; Sequence 16126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16126
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126

Query Match 5.2%; Score 126.8; DB 4; Length 2295;
Best Local Similarity 47.1%; Pred. No. 5.4e-16;
Matches 571; Conservative 0; Mismatches 592; Indels 49; Gaps 4;

2Y 516 CGGCTCGAGAACAGCTTCGCTCCGCTGAGCTGTCGCGCGCGCATGTCATCCGGCTCA 575
DB |||||
1168 CGGCTCGAGAACAGCTTCGCTCCGCTGAGCTGTCGCGCGCGCATGTCATCCGGCTCA 1109
2Y 576 CTGCTCAGCGCTGAGCACTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 635
DB |||||
1108 CAGCTGCGCGCGCGGCTTTCGCGCATCCGCGAGATGACACGCTGATCCGCT 1049
2Y 636 CTGGAACACCGCATCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 695
DB |||||
1048 GATCAACCGCAAGTCTATCCGACATCCGCTGAAGAGTGTGAGTGTGCTGCTGCTGCTG 989
2Y 696 CCGTACCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
DB |||||
988 CCGTACCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 929
2Y 756 CGTTTTCAGCAGGGAACCGAAGATCATGTTTGGCGCGAGGCTCATCGCTTTCG 815
DB |||||
928 CGGTG-----AGTGGCTGCCGCGCGCGCGAGAGCTGAGGCTGGGCGCG 887
2Y 816 TCTCGAGGCACTGCTGCTGCGCGCGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 875
DB |||||
886 GCTGAGCGCGCTGACCGCTGCGCGCGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 827
2Y 876 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 935
DB |||||
826 GTCCACCGCTTACCGCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 767
DB |||||

QY 936 GGCCTTGACGGCTCTCAGCGTGAAGCCATAGTGGCCAGAGGGCTCGTTGGCGCGTT 995
 Db 766 CGCTCGCGCGGGCTAGAGCTGAGGCCATGCTTCGGTTCCGGGGCGCCGTTGATGCGC 707
 QY 996 CATTCACGAGCTTCCGCGCCAGCCCGCCAGGTGAGGTCCGCGCAACATCCGAC 1055
 Db 706 CATTCATGCGCGCGCGCCAGCGTGGGAGATGACATGAGCGCGCGCTTATCCGACCT 647
 QY 1056 GCTCCTTTCCGGGTGCTCGTTTCCGTTGAGACAGAGAGAGGTCAAGGTCAAGAGCA 1115
 Db 646 GCTCACCGCCAGAGAGAGGTGAGCGCGCTCCCATAGAGTGCACAAAGT----- 596
 QY 1116 CGAGGCAATTTCCGCAAGACCGCTAACCGGCTCCGACGTCCGCTAGTTCCCTCGGCGC 1175
 Db 595 -----CGAGACCCCTATTCCGTGGTTGGCAGCCCGAGGTATGGCGC 551
 QY 1176 GCTCTGAGAGACATGATCAGCCTTACTCTGCTCTGCTCGAGAAACAACAGACGAC 1235
 Db 550 CTGCTGAGCCCGAGATGCGCCAGCGCGAGGTGTGAGATCGAAGCCAAACGCGTGTG 491
 QY 1236 CGACAAACCGGCTCTCTGAGAGTGAAGAAACAAGACCGCGCAACCGCGCAACTTCCAGCG 1295
 Db 490 CGACAAACCGGCTGTGATTTGCGCGCCAGAGGCGACGTGATCTCCGCGCAACTTCCAGCG 431
 QY 1296 GTCCGCTGTCTGATTTGATGAGAGAACAGAGCTCCACTCCCTCAATCCGACGCT 1355
 Db 430 CGAACCGGTGGAGATGGCGCGCGACACCTGCGCTGGCTGTGGCGAGATCGGTTGCG- 372
 QY 1356 CAACCTTCAAGCAGTACCGAGTGTCTCAAGCTGCAATGAACCGCGCGCTGCTTGTG 1415
 Db 371 -----TGTGGAACCGCGCATCTGCTATGATGAGACATGTCAGATTTGCGCGC 317
 QY 1416 CCTCGTGGCCAGAGACCGCGCTCACTATCAAGGAGGCTTGGACATTCACATGCG 1475
 Db 316 GTTCCGTGGGCGCAACCGCGGGGTCACTCCGCGTTGATGATCCCGAGGTACCGCGC 257
 QY 1476 TGTCTTACGCTTGGAGCTGCGCACCTTGCCACCGGCTCACTACTTGTCCAGCGCG 1535
 Db 256 CCGCGTGGCGCAAGCAACAGGCGCTGGCCATCCGCGCAAGCGCTGACGCTGCGGAC 197
 QY 1536 AAGATGGGCAACACAGCGCGCTCACTGCTGCTCATCTCCGCGCGCGCACTGCCA 1595
 Db 196 CTCGCGCAACACAGAGACCATGTGTGATGATGCGCCGCAACCGCGCAACCTGCGG 137
 QY 1596 GCGCAACGAGCTCTTTCTC-TCCTTTCGCTCGCACTGTACTGACGCTCCAGGCG 1654
 Db 136 GATGCGCGAAGACGTCCCGCGGATCTCGCGCTGAGTGGGTGGCGGCTGCGCAAGGCG 77
 QY 1655 TCGACCTCCGCGCGCATGAGCTCGACTTCAAGAGCATTCGACCGCGCTTCTCCGACTG 1714
 Db 76 TGGACTTCGCGAGGGGCTTGAAGATTGCGGAAGTGAAGACAGGCTCGCGCGCTCTGCT 17
 QY 1715 TCTTCACAGAGC 1726
 Db 16 GCGACAGGTGC 5

RESULT 15
 US-09-489-039A-6404
 ; Sequence 6404, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 6404

LENGTH: 1527
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-6404
 Query Match 4.5%; Score 108.4; DB 4; Length 1527;
 Best Local Similarity 48.2%; Pred. No. 1,9e-12;
 Matches 405; Conservative 0; Mismatches 396; Indels 39; Gaps 2;
 QY 517 GGCCTCGAACAACGCTTCGCTGAGGTGTCGCGCGCCAGGTATCCGCTCAAC 576
 Db 259 GGCCTTGGCGAGCGCGCTGACGATGATGCGCGCGCTGATTTAGTGTGAAGATTAC 318
 QY 577 TCGCTCAAGCTGGACACTCGCGCGCGCTGCTGTGCTTTAGAGCGCTCAACATTC 636
 Db 319 AGCTGTCCCGGTTTCTCGGAGTCCGCTTGAAGGTATCCAGCGCTGATTTGCGCTG 378
 QY 637 TTGAACACCGCATACGCCCATGCTCCCTCCGCGCTCATCTCGCGTCCGCGAC 696
 Db 379 GTGATGCGCGGTGTGACCGCGTGAATCCGCGCAAGAGCTCGTCCGCGCTTCCGCGAC 438
 QY 697 CTCAGCGCGCTCTGTACATCGCGCGCGCATACCGGTACCCGCACTCAAGTTAC 756
 Db 439 CTCGCGCGCTGGGCAATATGCTGACCTGATCGGAAGGTAAAGCGCGGTACCG 498
 QY 757 GTTTGACAGAGGAAACGAGAATCATGTTTCCGCGAGGCGCATCTCGCTTTGCT 816
 Db 499 GCGCGCGACTGCG-----TCCGCGCAACGAAAGCTGCTGATGAGTGGG 543
 QY 817 CTCAGGAGTGTGCTCTCGCGCGGAAAGAGGTGTGCTGTGATCAACGCGCGCTG 876
 Db 544 CTGAGGCGGATTAACCTCGCGCGGAAAGAGGCGCTGCGCTACTCAACGCGCACCGCGC 603
 QY 877 TCGCGCTGATGAGGACCTCAGTCTGACGAGTGGCAGATGGCATGTCCTCGCTCCGAC 936
 Db 604 TCGACCGCTTTTGCCTCGCGCGGTGTGTTGAAGCCGAGAGATTTGCTCGCTCCGCGCT 663
 QY 937 GCCTTGAAGGCTCTCAAGGTGAGAGCCATGCTGCGCGACAGAGGCTGTTCCGCGCTTC 996
 Db 664 GTCTGCGGGCGTTAACCAACGAGCGCGCTGCTCAAGTGTCTGTTATGATGCGGCG 723
 QY 997 ATCCAGACGTCTGCGCGCGCGCACCCCGCGCAAGTGAAGTGGCGCGCAACTCCGACG 1056
 Db 724 ATCCATGAAGTGGCGCGCGCGCACCGCGCAAGTGAAGTGGCGCGCTGATGCCACCTG 783
 QY 1057 CTCCTTCCGCGCTGCTGCTGTTGCGCTTGAAGACAGAGAGAGTCAAGGTCAAGAGCAG 1116
 Db 784 CTGACCGACGATAGCGGATCTGCGAGTGCACATTAATCGACAGAGGT----- 833
 QY 1117 GAGGGCAATTTTGGCCAGAGACCGCTAACCGGCTCCGACGTCGCTCACTTCTCGGCGC 1176
 Db 834 -----GAGGACCCGTTACTCTCCCTGCGCTGCGACCGCGAGGTATGGGCGC 879
 QY 1177 CTGCTGAGAGACATGATGACAGCGCTACTGACATCTCTGCTGAGAAACAACGACGAC 1236
 Db 880 TGCCTGACGAGATCCGCCAGCGCGCTGAGTGTGCTGGCTGAAGCCAAACCGGTTCTC 939
 QY 1237 GACAAACCGGCTCTCTGACGTCGAGAACAGACAGCGCGCACGCGCGCACTTCCAGCG 1296
 Db 940 GATTAACCGGCTGTCTTTGGCGCGGAAATGACGTATCTCCGGGAGTAATCTCATGCC 999
 QY 1297 TCGGCTGTCTGATTTGATGAGAAAGACAGGCTCGCACTGGCGCTCATCGGCAAGCT 1356
 Db 1000 GAAACCGGTGGAGTGGCGCGGATATATGCGGCTGGCGGATCGCGAATGCGGCTGCTC 1059

Search completed: September 12, 2004, 04:26:43
 Job time : 190.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M nucleic - nucleic search, using sw model

Run on: September 11, 2004, 23:12:54 (without alignments)
7750.794 Million cell updates/sec

Title: US-09-939-408A-12

Perfect score: 2419
1 ctctcgtctactactaacc.....gagttcaaaaaaaaaa 2419

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3304383 segs, 2515761380 residues

Total number of hits satisfying chosen parameters: 6608766

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10C_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US00_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US00_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2418.6	100.0	2419	9	US-09-939-408A-12
2	1826.6	75.5	2741	9	US-09-939-408A-28
3	1704.4	70.4	2163	9	US-09-939-408A-29
4	1629	67.3	2475	9	US-09-939-408A-20
5	1196.6	49.5	2151	9	US-09-765-873A-9
6	1196.6	49.5	2151	15	US-10-138-970A-23
7	1196.6	49.5	2151	15	US-10-188-523B-9
8	1196.6	49.5	2151	16	US-10-374-366-3
9	1196.6	49.5	2439	9	US-09-939-408A-18
10	1195	49.4	2151	9	US-09-765-873A-7
11	1195	49.4	2151	15	US-10-188-523B-7
12	1195	49.4	2151	16	US-10-374-366-1
13	1195	49.4	2151	16	US-10-374-366-9
14	1195	49.4	2151	16	US-10-374-366-13

15	1195	49.4	2151	16	US-10-374-366-23	Sequence 23, Appl
16	1193.4	49.3	2151	16	US-10-374-366-5	Sequence 5, Appl1
17	1193.4	49.3	2151	16	US-10-374-366-7	Sequence 7, Appl1
18	1193.4	49.3	2151	16	US-10-374-366-15	Sequence 15, Appl1
19	1193.4	49.3	2151	16	US-10-374-366-21	Sequence 21, Appl1
20	1191.8	49.2	2151	16	US-10-374-366-11	Sequence 11, Appl1
21	1190.2	49.2	2151	16	US-10-374-366-17	Sequence 17, Appl1
22	1190.2	49.2	2151	16	US-10-374-366-19	Sequence 19, Appl1
23	1188.8	49.1	2061	9	US-09-765-873A-31	Sequence 31, Appl1
24	1188.8	49.1	2061	15	US-10-188-523B-31	Sequence 31, Appl1
25	1186.6	47.0	2787	9	US-09-939-408A-16	Sequence 16, Appl1
26	802.4	33.2	2703	16	US-10-439-478-1	Sequence 1, Appl1
27	696.8	28.8	2465	17	US-10-439-478-24	Sequence 24, Appl1
28	545	22.5	1020	15	US-10-138-970A-3	Sequence 3, Appl1
29	424.8	17.6	2554	17	US-10-767-701-15815	Sequence 15815, A
30	423.2	17.5	2539	13	US-10-425-114-26376	Sequence 26376, A
31	408.2	16.9	2325	17	US-10-767-701-15813	Sequence 15813, A
32	402	16.6	2544	17	US-10-437-963-5561	Sequence 5561, A
33	401.6	16.6	2454	13	US-10-425-114-34124	Sequence 34124, A
34	400.2	16.5	2417	13	US-10-425-114-31197	Sequence 31197, A
35	400.2	16.5	2468	13	US-10-425-114-30888	Sequence 30888, A
36	400.2	16.5	2469	13	US-10-425-114-31903	Sequence 31903, A
37	400	16.5	2354	13	US-10-425-114-28070	Sequence 28070, A
38	385.8	15.9	2972	17	US-10-437-963-48311	Sequence 48311, A
39	383	15.8	2774	17	US-10-437-963-38582	Sequence 38582, A
40	382.8	15.8	2329	17	US-10-437-963-41428	Sequence 41428, A
41	382.2	15.8	2079	13	US-10-425-114-1334	Sequence 1334, A
42	380.4	15.7	1092	17	US-10-439-479-9	Sequence 9, Appl1
43	376.2	15.6	2340	13	US-10-425-114-2857	Sequence 2857, Ap
44	350.8	14.5	2145	17	US-10-437-963-40094	Sequence 40094, A
45	346	14.3	2535	17	US-10-437-963-36111	Sequence 36111, A

ALIGNMENTS

RESULT 1
US-09-939-408A-12
Sequence 12, Application US/09939408A
Patent No. US20020102712A1
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500SCA
CURRENT APPLICATION NUMBER: US/09/939, 408A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0.
SEQ ID NO 12
LENGTH: 2419
TYPE: DNA
ORGANISM: Rhodotorula graminis
FEATURE:
NAME/KEY: CDS
LOCATION: (37)..(2196)
NAME/KEY: modified_base
LOCATION: (494)
OTHER INFORMATION: Other information: y = t or c
NAME/KEY: unsure
LOCATION: (493)..(495)
OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408A-12
Query Match 100.0%, Score 2418.6, DB 9, Length 2419,
Best Local Similarity 100.0%, Pred. No. 0,
Matches 2419, Conservative 0, Mismatches 0, Indels 0, Gaps 0,

QY	1	CTCTCGCTCACTCTAAACCCGCTACTAGCACTTGCATGAGCCCTTCTCTTGAATCGCTC	60
Db	1	CTCTCGCTCACTCTAAACCCGCTCACTAGCACTTGCATGAGCCCTTCTCTTGAATCGCTC	60
QY	61	GCCACCAAGCTCGCCCAAGGCTTACCAAGGCTGCAGCGGCTCGACCAAGTGGAGCT	120
Db	61	GCCACCAAGCTCGCCCAAGGCTTACCAAGGCTGCAGCGGCTCGACCAAGTGGAGCT	120
QY	121	GCGGAGCCCACTTGGGCTTCCGCGCAGCCCGGCTTGATGGCAGCCGCGACAG	180
Db	121	GCGGAGCCCACTTGGGCTTCCGCGCAGCCCGGCTTGATGGCAGCCGCGCGACAG	180
QY	181	TGCAAGCTCGAGATCGTGCAGAGAGCTCTCAGAGACCCACCGAGAGAGTGTGAGCTC	240
Db	181	TGCAAGCTCGAGATCGTGCAGAGAGCTCTCAGAGACCCACCGAGAGAGTGTGAGCTC	240
QY	241	AGCGGGTACAGCCTCACCGTCCGTGACGTGTGCGCGCGCCGCAAGAGGCGCAGGCTC	300
Db	241	AGCGGGTACAGCCTCACCGTCCGTGACGTGTGCGCGCGCCGCAAGAGGCGCAGGCTC	300
QY	301	GCGCTCCAGAACCAACGACGAGATCCCGCGACCGCTGCACAGAGCTGACTTCTTAAG	360
Db	301	GCGCTCCAGAACCAACGACGAGATCCCGCGACCGCTGCACAGAGCTGACTTCTTAAG	360
QY	361	GCCCAAGCTTCAAGACTCGGCTACAGAGTACCAAGGATTCCGGGTGCGGCGGACAG	420
Db	361	GCCCAAGCTTCAAGACTCGGCTACAGAGTACCAAGGATTCCGGGTGCGGCGGACAG	420
QY	421	AGACTAGAGATGACGTCAAGCTTCCAGAAAGGCGCTCATGAGACCAAGCTTGGGCGTG	480
Db	421	AGACTAGAGATGACGTCAAGCTTCCAGAAAGGCGCTCATGAGACCAAGCTTGGGCGTG	480
QY	481	ACGCGCAAGTCCGCTCGTCTTACGCGCGTCCGAGCCCGGCTTCGAGAACAGGCTTCCGCTC	540
Db	481	ACGCGCAAGTCCGCTCGTCTTACGCGCGTCCGAGCCCGGCTTCGAGAACAGGCTTCCGCTC	540
QY	541	GAGGTGCTCGCGCGCCCAATGATCATCCGCGTCAAGCTGCACAGCTGCGCACCTCGAGC	600
Db	541	GAGGTGCTCGCGCGCCCAATGATCATCCGCGTCAAGCTGCACAGCTGCGCACCTCGAGC	600
QY	601	GTCGCGCTCGTCTCTTGAAGCGCTCACCACTTGTGAACACCGCATCACGCCATC	660
Db	601	GTCGCGCTCGTCTCTTGAAGCGCTCACCACTTGTGAACACCGCATCACGCCATC	660
QY	661	GTCGCGCTCGCGGCTCCATCTCGGCGTCCGCGACCTCAGCCCGCTCTGTGACATCGCC	720
Db	661	GTCGCGCTCGCGGCTCCATCTCGGCGTCCGCGACCTCAGCCCGCTCTGTGACATCGCC	720
QY	721	GCGGCACTCACCGGTCAACCCGAGCTCAAGGTTCAAGTTTGCACAGAGGAAACGAGAG	780
Db	721	GCGGCACTCACCGGTCAACCCGAGCTCAAGGTTCAAGTTTGCACAGAGGAAACGAGAG	780
QY	781	ATCATGTTTACGCGGAGGCAATCTCGCTTTTGTGTGTGAGAGAGTGTCTCGGCGCG	840
Db	781	ATCATGTTTACGCGGAGGCAATCTCGCTTTTGTGTGTGAGAGAGTGTCTCGGCGCG	840
QY	841	AAGAGAGGCTTCGCTGTGTCACAGGAACGAGCGCTTCGCTGATGAGCAACCTCAAT	900
Db	841	AAGAGAGGCTTCGCTGTGTCACAGGAACGAGCGCTTCGCTGATGAGCAACCTCAAT	900
QY	901	CTGCAAGACTCCGACATGCTCTCGCTCTCTTCGAGAGCTTGAAGGCTTTCACAGGTGAG	960
Db	901	CTGCAAGACTCCGACATGCTCTCGCTCTCTTCGAGAGCTTGAAGGCTTTCACAGGTGAG	960
QY	961	GCCATGCTCGGACAGAGGCTCTTTCGCGCCGTTTCAACACAGCTTGCAGCCCGAC	1020
Db	961	GCCATGCTCGGACAGAGGCTCTTTCGCGCCGTTTCAACACAGCTTGCAGCCCGCGAC	1020
QY	1021	CCCGGCAAGTGAAGTGCAGGCAACTCCGACGCTCTTTCGCGCTGTGTGCTTGC	1080
Db	1021	CCCGGCAAGTGAAGTGCAGGCAACTCCGACGCTCTTTCGCGCTGTGTGCTTGC	1080

QY	1091	GTTSAGACAGAGAGAGAGGTCAAGGTCAAGACACAGAGGCATTCTTCGCGAGACCGC	114
Db	1081	GTTAGAGACAGAGAGAGAGGTCAAGGTCAAGACACAGAGGCATTCTTCGCGAGACCGC	114
QY	1141	TACCGCTCCGACAGTCGTGAGTTCCTCGGCGCGCTCGGAGAGCATGATGCAGCC	120
Db	1141	TACCGCTCCGACAGTCGTGAGTTCCTCGGCGCGCTCGGAGAGCATGATGCAGCC	120
QY	1201	TACTTCGACTCTCTCGCTCGAAGAACACACACCGAACACCCGCTCTCGAGCTCGAG	126
Db	1201	TACTTCGACTCTCTCGCTCGAAGAACACACACCGAACACCCGCTCTCGAGCTCGAG	126
QY	1261	AACAAGACAGACCGCGACCGCGGGCAACTTCCAGCGTCGTGGGTGCTGATTTGATGGAG	132
Db	1261	AACAAGACAGACCGCGACCGCGGGCAACTTCCAGCGTCGTGGGTGCTGATTTGATGGAG	132
QY	1321	AAGACAGAGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCAACGAGTGCACCGAGTTG	138
Db	1321	AAGACAGAGCTCGCACTCGCCCTCATCGGCAAGCTCAACTTCAACGAGTGCACCGAGTTG	138
QY	1381	CTCAACGCTGCCATGAACCGGCGCGCTGCTTGTCGCTCGCTCGACGAGAACCCGTCGTC	144
Db	1381	CTCAACGCTGCCATGAACCGGCGCGCTGCTTGTCGCTCGCTCGACGAGAACCCGTCGTC	144
QY	1441	AACATACAGGCAAGGCGTTGGACATTACATCGCTCTTACGCTTCGAGAGCTCGAGCAC	150
Db	1441	AACATACAGGCAAGGCGTTGGACATTACATCGCTCTTACGCTTCGAGAGCTCGAGCAC	150
QY	1501	CTTGCCCAACCCCGTCACTACCTTGTCGTCAGCCCGCAGACATGGGCAACAGCCGCTCAAC	156
Db	1501	CTTGCCCAACCCCGTCACTACCTTGTCGTCAGCCCGCAGACATGGGCAACAGCCGCTCAAC	156
QY	1561	TGCGCTCGCTCATCTCGCGCGCGCGCACACACGAGGCAACAGACGCTCTTCTCTCGCTT	162
Db	1561	TGCGCTCGCTCATCTCGCGCGCGCGCACACACGAGGCAACAGACGCTCTTCTCTCGCTT	162
QY	1621	CTTGCGCTCGCACCTGTACTGACGCTCOAGGCGGTGACCTTCGCGCGATGGAGCTCGAC	168
Db	1621	CTTGCGCTCGCACCTGTACTGACGCTCOAGGCGGTGACCTTCGCGCGATGGAGCTCGAC	168
QY	1681	TTCAAGAACAGATTGACACCGCTTCTCCGACCTCTCTCAGAGAGACCTCGGCACTGGC	174
Db	1681	TTCAAGAACAGATTGACACCGCTTCTCCGACCTCTCTCAGAGAGACCTCGGCACTGGC	174
QY	1741	CTTGACGTCGAACGCACTTTCGCTTGAGGTCAAGAGCGCTCAACAGCGTTCGACGAG	180
Db	1741	CTTGACGTCGAACGCACTTTCGCTTGAGGTCAAGAGCGCTCAACAGCGTTCGACGAG	180
QY	1801	ACGACGAGTGAAGACTCTGAGACCGCGCTGGCAAGAGCCTTCTGTAAGCAACCGGACCC	186
Db	1801	ACGACGAGTGAAGACTCTGAGACCGCGCTGGCAAGAGCCTTCTGTAAGCAACCGGACCC	186
QY	1861	GTTCGTGAGCTCTCTCGCTCTCGCCCTTGCCAAAGTCAACCTTACTAGCGGTCAACGCG	192
Db	1861	GTTCGTGAGCTCTCTCGCTCTCGCCCTTGCCAAAGTCAACCTTACTAGCGGTCAACGCG	192
QY	1921	TGGAAGGTTGCTCGCGCTCGAGAAAGGCAATGTGCTTACGCGGAGGTGCGAACCGCTTC	198
Db	1921	TGGAAGGTTGCTCGCGCTCGAGAAAGGCAATGTGCTTACGCGGAGGTGCGAACCGCTTC	198
QY	1981	TGGCAGACGCGGTCTTGCGACAGCGCGCGCGCAACGCACTCTTCGCGCGACGCGCTC	204
Db	1981	TGGCAGACGCGGTCTTGCGACAGCGCGCGCGCAACGCACTCTTCGCGCGACGCGCTC	204
QY	2041	CTGTACTCTGTTCTGTGCGCGAGAGCTTGCGGTGCAAGCGCGCGCGGAGCATGTATTGTC	210
Db	2041	CTGTACTCTGTTCTGTGCGCGAGAGCTTGCGGTGCAAGCGCGCGCGGAGCATGTATTGTC	210
QY	2101	GCGCGAGACAGAGAGAGATCGGAGACAAAGTCTGCGCATTTAGAGGCCATTAAGAAC	216
Db	2101	GCGCGAGACAGAGAGAGATCGGAGACAAAGTCTGCGCATTTAGAGGCCATTAAGAAC	216
QY	2161	GCGCGCATCAACAGCTCTGTCAAGATGCTTCGTGAAGCGCGAGCAACCTCGCTA	222

Db 2161 GGGCCCATCAACCAAGTCTCGTCAAGATGCTCCGCGTAAGCCCGACAGCCTCGCTTA 2220
 Oy 2221 GAGCGCCGCTCAGACCCCAAGACAGCTTTTGAAGTGTGTCGACCAAGAGCACTTT 2280
 Db 2221 GAGCGCCGCTCAGACCCCAAGACAGCTTTTGAAGTGTGTCGACCAAGAGCACTTT 2280
 Oy 2281 CCTCATACACATGTGCGCTTACTCTCTGCGCTCATCAGTCTCTCATGTTCTTTGGTAT 2340
 Db 2281 CCTCATACACATGTGCGCTTACTCTCTGCGCTCATCAGTCTCTCATGTTCTTTGGTAT 2340
 Oy 2341 CCGCGGTCTCTGCGTCAGTACACCGGTATAGAGCGCTGGAATGATGCAAGTCTTG 2400
 Db 2341 CCGCGGTCTCTGCGTCAGTACACCGGTATAGAGCGCTGGAATGATGCAAGTCTTG 2400
 Oy 2401 AGTTCAAAAA 2419
 Db 2401 AGTTCAAAAA 2419

RESULT 2

US-09-939-408a-28
 Sequence 28, Application US/09939408A
 Patent No. US20020102712A1

GENERAL INFORMATION:

APPLICANT: Yoshida, Roberta
 APPLICANT: Koostra, Anna
 TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 TITLE OF INVENTION: Using Same
 FILE REFERENCE: 29479/500NSCA
 CURRENT APPLICATION NUMBER: US/09/939,408A
 CURRENT FILING DATE: 2001-08-24
 PRIOR APPLICATION NUMBER: US 09/624,693
 PRIOR FILING DATE: 2000-07-24
 PRIOR APPLICATION NUMBER: PCT/US01/23270
 PRIOR FILING DATE: 2001-07-24
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 28
 LENGTH: 2741

TYPE: DNA

ORGANISM: Rhodocycla graminis

FEATURE:

NAME/KEY: modified_base

LOCATION: (6) OTHER INFORMATION: Description of modified_base: m = a or c

NAME/KEY: exon

LOCATION: (2008)..(2586)

OTHER INFORMATION:

NAME/KEY: exon

LOCATION: (1822)..(1947)

OTHER INFORMATION:

NAME/KEY: exon

LOCATION: (1587)..(1748)

OTHER INFORMATION:

NAME/KEY: exon

LOCATION: (1365)..(1529)

OTHER INFORMATION:

NAME/KEY: exon

LOCATION: (961)..(1295)

OTHER INFORMATION:

NAME/KEY: exon

LOCATION: (449)..(880)

OTHER INFORMATION:

NAME/KEY: exon

LOCATION: (1)..(361)

OTHER INFORMATION:

NAME/KEY: terminator

LOCATION: (2587)..(2589)

OTHER INFORMATION:

NAME/KEY: Intron

LOCATION: (1948)..(2007)

OTHER INFORMATION:
 NAME/KEY: Intron
 LOCATION: (1749)..(1821)
 OTHER INFORMATION:
 NAME/KEY: Intron
 LOCATION: (1530)..(1586)
 OTHER INFORMATION:
 NAME/KEY: Intron
 LOCATION: (1296)..(1364)
 OTHER INFORMATION:
 NAME/KEY: Intron
 LOCATION: (881)..(960)
 OTHER INFORMATION:
 NAME/KEY: Intron
 LOCATION: (362)..(448)
 OTHER INFORMATION:
 US-09-939-408a-28

Query Match 75.5%, Score 1826.6, DB 9, Length 2741;
 Best Local Similarity 84.3%, Pred. No. 0;
 Matches 2312; Conservative 2; Mismatches 1; Indels 426; Gaps 6;

Oy 37 ATGCGCCCTTCTCTGAGTGTGCTGCAACCAAGCTGCGCAAGGCTTTTACCAACGCTCG 96
 Db 1 ATGCGMCTTCTCTGAGTGTGCTGCAACCAAGCTGCGCAAGGCTTTTACCAACGCTCG 60
 Oy 97 CAGCGCGCTCGCAACCAAGTGGGCTGCGGGGCCCACTTGCGCTCGCGCGCAGCGCGG 156
 Db 61 CAGCGCGCTCGCAACCAAGTGGGCTGCGGGGCCCACTTGCGCTCGCGCGCAGCGCGG 120
 Oy 157 CTGATGCGCAAGCGCGCGCAACCAAGTGGGCTGCGGGGCCCACTTGCGCTCGCGG 216
 Db 121 CTGATGCGCAAGCGCGCGCAACCAAGTGGGCTGCGGGGCCCACTTGCGCTCGCGG 180
 Oy 217 CCCACCGCAAGCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 276
 Db 181 CCCACCGCAAGCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 240
 Oy 277 GCGCGCGCGCAAGGCGCGAGGCTGCGGCTCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 336
 Db 241 GCGCGCGCGCAAGGCGCGAGGCTGCGGCTCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 300
 Oy 337 GACCAAGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 395
 Db 301 GACCAAGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 360
 Oy 396 ----- 395
 Db 361 GTGCGTCCGAGACGAGAGGCGGAAATCTCGGAGTGCCTGAGCGCTGAACGCTGACATC 420
 Oy 396 ----- 429
 Db 421 GTTGAAGGCTGCGCGGCTGCTTGAAGGCTTTCGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 480
 Oy 430 GATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 489
 Db 481 GATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 540
 Oy 490 TCCGCTGCTGCTTCAAGGCTGAGAGCGCGGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 549
 Db 541 TCCGCTGCTGCTTCAAGGCTGAGAGCGCGGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 600
 Oy 550 CCGCGCGCGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 609
 Db 601 CCGCGCGCGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 660
 Oy 610 GTGCTGCTTGAAGGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 669
 Db 661 GTGCTGCTTGAAGGCTGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 720
 Oy 670 CCGCGCTCATCTCGAGCTGTGAGGCGCACTTCAAGCGCTGTGATCATGTGCGGCGCATC 729
 Db 721 CCGCGCTCATCTCGAGCTGTGAGGCGCACTTCAAGCGCTGTGATCATGTGCGGCGCATC 780

Db 1261 TTTCAGGCGGCGCTGTGCGAACAACATGAGAGAAAGCTGCGCTCCGCTCCGCTCGATC 1320
2y 1348 GGCAAGCTCAACTTCAGCAGAGTGCACCGAGTTGTCAACGCTGCATGAACCGCGCGCTG 1407
Db 1321 GGCAAGCTCAACTTCAGCAGAGTGCACCGAGTTGTCAACGCTGCATGAACCGCGCGCTG 1380
2y 1408 CTTTGTGCTCGTGTGCGGAGAGCCGCTGCTCAACTATCAGCGCAAGGGCTTGAATT 1467
2y 1381 CCGTCTGCTGCTGCTGCGAGAGACCGTCTGCTCTATCATCTGCAAGGGCGCTCGAATT 1440
2y 1468 CACATCGTGTTAAGCTTGGAGTCTGCGCACTGCGCAACCGGTCCTACTCTGCTGCTG 1527
2y 1441 GCGCGCGCGCTGCTGAGCTGCGACCTTGTCCAACTCGGTCGAGCCGCTGCTGCTGCTG 1500
2y 1528 CAGCGCGCGAGAGTGGGCAACAGCGCGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTG 1587
2y 1501 CAGCGCGCGAGAGTGGGCAACAGCGCGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
2y 1588 ACTGCGGAGGCGCAACAGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1647
2y 1561 ACNCGCGAGGCGCAACAGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
2y 1648 CAGCGCGCGAGAGTGGGCAACAGCGCGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTG 1707
2y 1621 CAGCGCGCGAGAGTGGGCAACAGCGCGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
2y 1708 CCGACTCTCTCCAGCAGACCTCGGCACTGCGCTCGAGCTC-----AACGA 1755
2y 1681 CCGACTCTCTCCAGCAGACCTCGGCACTGCGCTCGAGCTC-----AACGA 1740
2y 1756 CTTGCGCTTCAGAGTCAAGAAAGCGCTCAACAGAGTTCAGAGAGTTCAGAGAGTTCAGAG 1815
2y 1741 CTTGCGAGCAAGGTTCAAGAAAGCGCTCAACAGAGTTCAGAGAGTTCAGAGAGTTCAGAG 1800
2y 1816 CTTGAGCGCGCGTGCAGAGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1875
2y 1801 CTTGAGCGCGCGTGCAGAGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
2y 1876 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1933
2y 1861 NNCTCTCTGCT 1920
2y 1934 CCGCGCGAGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1993
2y 1921 CCGCGCGAGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
2y 1994 CTTGCGAGGCTGCT 2053
2y 1981 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
2y 2054 TCGCGAGAGAGTTCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 2113
2y 2041 TCGCGAGAGAGTTCGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 2100
2y 2114 ACAGCATGGGAGCAAGCTGCTGCGCATCTGAGAGGCGCATCAAGAGCGCGCGCTCAAC 2173
2y 2101 TGAAGATGGGAGCAAGCTGCTGCGCATCTGAGAGGCGCATCAAGAGCGCGCGCTCAAC 2160
2y 2174 ACCTCTCTGCTGAAGTGTGCGCTGTA-AGCGCGAGCAAGCGCTGCTGAGCGCGCGCTG 2232
2y 2161 ACCTCTCTGCTGAAGTGTGCGCTGTA-AGCGCGAGCAAGCGCTGCTGAGCGCGCGCTG 2220
2y 2233 ACCCGAAGCAAGCTTTCAGAGTGTGCTGCTGCGCAAGAGCGCTTCTGCTGCTGCTGCTG 2292
2y 2221 NNCCNN 2280
2y 2293 TGTGCGCTTACTCTGCGCGCTGATCAAGTCTCTGAGTTCTTCTGTAATCCGCGCTCTCTC 2352
2y 2281 TNN 2340
2y 2353 GGTGCTCA-GTACAGGTATAGAGCTGGAATGATTTGCAAGTCTTGAATTAATAA 2411

Db 2341 NGTNNCANNACCTNN 2400
2y 2412 AAAAAAAAA 2419
Db 2401 NNNNAAAA 2408

RESULT 5
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: ECI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-9

Query Match 49.5%; Score 1196.6; DB 9; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2,4e-282;
Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

2y 99 CGCGCTTCGACCAAGTGTGCGTGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCT 158
2y 45 CGTGCATCCGCAAGAGAGGCTGTCATGCGCGCTGCAACCACTGCGAGTCCAGAGCTTC 104
2y 159 CGATGGCCACCGCT 218
2y 105 GCACCTGCGCAACCGAGGTGACGAGTGCATGTCGAGAGAGTGTCTGCGCGCGCT 164
2y 219 CACCGAGAGCTGCTGCGAGCTGACGCGGCTACAGCTCACCGCTGCTGCTGCTGCTGCTG 278
2y 165 GACGATCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTG 224
2y 279 CGCGCGCAAGGCGCGAGGCTGCGCTGCGCAAGAGCGAGCGAGTCCGCTCAAGATTGA 284
2y 225 CGCGAGAGAGGCGAGGCTGCGCTGCGCAAGAGCGAGCGAGTCCGCTCAAGATTGA 284
2y 339 CAGAGAGCTGCACTTCTCAAGGCGCGCGCTTCAAGACTGCGTCTAGAGGCTCAACGCG 398
2y 285 CAATGAGTGCAGTCTTGTGCGCTGCAACTCTCTCAAGAGTCTTCAAGAGTCTTCAAGAG 344
2y 399 TTTGAGTGTGCGCGCGAGCAAGAGTGTGAGATGAGTCAAGCTTCCAGAGGCGCTGAT 458
2y 345 ATTGCGGATTCGAGACACCGCGACCGAGAGCGCATGCGCTCAGAGAGGCTCTCT 404
2y 459 CGAGACCAAGCTTTCGCGCGTGAAGCGCGAGTCCGCTGCTGCTGCTGCTGCTGCTGCTG 518
2y 405 CGAGACCAAGCTTTCGCGCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 464
2y 519 CCGTGAAGACAGGCTTTCGCTGAGAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCTCACTC 578
2y 465 TTTGAGAGTGTGCTTCTGCTGAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCTCACTC 524
2y 579 GCTCAAGGCGCGCACTGCGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
2y 525 CTTGACCGCGCGCACTGCGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
2y 639 GAACACCGCATCAAGCGCGCATGCTGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCT 698

Db 585 CAACCAAGGATACCCCAATGTCCTCCCGGACCACTCTGTGGGCGAGCT 644
 Qy 699 CAGCCGCTCTCTATACATGCGCGCCATCAACCGGTACCCCGCATCAAGTTCAAGT 758
 Db 645 CTCTCTCTCTCTATATGCGCGCCATCAAGCGGTACCCCGACAGAGGTGACGT 704
 Qy 759 TTTCGACGAGGGAACCGAAGATATATGTTGCGCGGAGGCGCATCTGCTTTGGTCT 818
 Db 705 CGTCCACGAGGGAAGAGATCTGTACCGCCGCGAGGAGATGCGCTTTCAACT 764
 Qy 819 CGAGGAGTGTCTTGGCGCGGAGGAGGTCTGGTGTGTCACAGGAAAGGCGGTCTC 878
 Db 765 CGAGCCGCTGTCTCTGCGCCGAGAGGAGGTCTGAGTCTGTCAACGAGCAGCGCTCTC 824
 Qy 879 CGCTGTATGCGGACCTCTCAATCTGACAGCTGCGCATGCTCTGCTCTCTCGAGGC 938
 Db 825 AGCATGATGCGCACCTCGCTCTGCAAGAGCAACATGCTCTGCTCTCTCGAGTC 884
 Qy 939 CTTCAGGCTCTCAAGTGGAGGAGCATGATGCGGCTGAGAGGCTTCTGCGCGCTCAT 998
 Db 885 GCTCAAGGCAATGACGCTTGAAAGCATGAGTGGCCACGCGGCTGTTCACACCTTCTCT 944
 Qy 999 CCAAGAGTCTGCGCGCGGACCCCGGACAGGTGAGGTGCGGCGCAACATCCGACGCT 1058
 Db 945 TCACGAGCTCAAGCGCCCTCACCCGACGAGATGCAATCGCGGAGAAACATCCGACGCT 1004
 Qy 1059 CCTTTCGCGCTGTGTGTTGCGGTGAGCAAGAGAGAGAGTCAAGTCAAGACGACGA 1118
 Db 1005 CCTCGAGGGAAGCCGCTTGTGTGTCTCACATGAGAGAGAGTCAAGTCAAGACGACGA 1064
 Qy 1119 GGGCATTTCTTGGCGAGACCGGCTACCGGCTCGGACGCTGCTTAACTTCTCGGCGCT 1178
 Db 1065 GGGCATTTCTTGGCGAGACCGGCTTACCTTGGGACGCTCTTCTAGTGGCTCGGCGCT 1124
 Qy 1179 CGTGAAGACATGATGACAGCGCTACTCGACTCTCTGCTCGAGAAACA---GACGAC 1235
 Db 1125 CGTGAAGACCTCATTCAGGCGCCAGCGCTCTCATCATGAGGCGGCGAGTGAAGAC 1184
 Qy 1236 CGACAAACCGCTCTCTCGAGCTGAGAAACAGCAACCGCGAGCGGCGCAACTTCCAGGC 1295
 Db 1185 CGACAAACCTCTCATGAGCTGAGAAACAGCAACTTCCACCGCGCGCAATTTCCAGGC 1244
 Qy 1296 GTGCGCTGTCTGATTTGATGAGAAAGACAGGCTGCACTGCGCTCTCAAGGAGCT 1355
 Db 1245 TGGCGCTGTGGCCAAACATGAGAAAGCTCGCTCGGCTCGCCAGATCCGACGCT 1304
 Qy 1356 CAATTCAGGAGTGCACCGAGTTGCTCAACGCTGCCATGAACCGCGGCTGCTTCTGTG 1415
 Db 1305 CAATTCAGGAGTGCACCGAGTGTCAACGCTGCCATGAACCGCGGCTTCCCTCTGTG 1364
 Qy 1416 CTTGGCTGCGAGGACCGGCTGCTCAACTATCAAGGAGGCTTGAACATTCACATGCG 1475
 Db 1365 CTTGGCGCTGAGAGGAGGCTGCTCTCTTACATGAGAGGCTTGCACATGCGCGCTGC 1424
 Qy 1476 TGTTCAGCTTTCGAGCTGCGCACTTGGCAACCGGCTCACTACTTCTGCGAGCGGC 1535
 Db 1425 GGGCTACACCTGAGGAGTGGAGACCTGCGCAACCTGTACAGAGATGTCCAGCGGC 1484
 Qy 1536 AGAGATGAGCAACCGAGGCTGCACTGCTGCTTATCTTCCGCGGCGCGCATCTCGGA 1595
 Db 1485 TGAAGTGGGGAACGAGGAGTCACTGCTTGGCTCATCTGCTGCTGCTGCGACGACGA 1544
 Qy 1596 GGGCAACGAGCTCTTCTCTCTCTCTGCGCTGCACTGTATGACGCTCGAGCGCT 1655
 Db 1545 GTTCAAGAGAGCTCTTCTCTCTCTCTGCGCAACCTTACTGTGCTTCTCAAGCT 1604
 Qy 1656 CGACTTGGCGGAGTGAAGCTGCACTTCAAGAAAGAGTGAAGCTTCTTCCAGCTCT 1715
 Db 1605 CGACTTGGCGGAGTGAAGTGAAGTGAAGAGAGTGAAGAGTGAAGAGTCTGCT 1664
 Qy 1716 CTTCAAGAGAGCTTGGCACTGAGCTGCAAGTCA-----ACGCACTTGGCT 1763

Db 1665 CATGACAGCACTTTGCTCCGCAATGACGAGCTCGAAGCTTGGCGAGAGCTGTGCA 1724
 Qy 1764 CGAGTCAAGAGGCGCTCAACAGGCTTCTGAGCAGACGACGATGACCTCGAGGC 1823
 Db 1725 GAAAGTGAACAGAGCTTCTGCGCAAGCGCTTGAACACCAACTGTGACATCTTCTCT 1784
 Qy 1824 GCGCTGAGCAGACGCTTCTTGTGAAGAGCGGACCGGCTGTGAGCTCTCTCTCTCTC 1883
 Db 1785 GCGCTGAGCAGACGCTTCTCTCTTCTGCGCGGACCGTGTGAGAGTCTCTCTCTCTC 1844
 Qy 1884 GCGCTTCTCAACGCTCAACCTTACTGCTTCAAGCTTGAAGTGTGCTTCTGCGCAGAA 1943
 Db 1845 GTC-----GCTCTGCTGCGCGCTGCAAGCTTGAAGTGTGCGCGCGCGCAAGTC 1895
 Qy 1944 GCGCATCTGCTCAAGCGCGGAGGTGCGCAACCGCTTCTGCAAGCGCGCTTCTGCGAGGC 2003
 Db 1896 GCGCATCTGCTCAAGCGCGGAGGTGCGCAACCGCTTCTGCAAGCGCGCTTCTGCGAGGC 1955
 Qy 2004 GCGCGCGCAGCATATCTTCTGCGCGCAGCGGCTGTGATCTGTTGTGCGCAGGA 2063
 Db 1956 GCGCGCGCTCTGTACTCTTCTGCGCGCAGCATGATCTTCTGAGCTTCTGCGCAGGA 2015
 Qy 2064 GCTGCGGCTGAGGCGCGCGCGCGCGCGAGCTTGTGCGGCTGCGCAGGAGAGATGCG 2123
 Db 2016 GCTTGGCTCAAGCGCGCGCGCGCGCGAGCTTCTGCGGAGAGAGAGTGAAGATGCG 2075
 Qy 2124 GAGCAGCTCTGCGCATCTTACGAGGCTCAAGAGCGCGCGCATCAACAGCTCTCTCT 2183
 Db 2076 CTCGACGCTTCCAAATCTTACGAGGCGCATCAAGGAGAGATCAACAGCTCTCTCT 2135
 Qy 2184 CAAGATGCTGCGCTA 2198
 Db 2136 CAAGATGCTGCTTA 2150

RESULT 6
 US-10-138-970A-23
 ; Sequence 23, Application US/10138970A
 ; Publication No. US2003079255A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sarsiliani, Sima
 ; APPLICANT: Tang, Xiao-Song
 ; APPLICANT: Qi, Wei Wei
 ; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
 ; TITLE OF INVENTION: hydroxycinnamic Acid
 ; FILE REFERENCE: CUI777
 ; CURRENT APPLICATION NUMBER: US/10/138,970A
 ; CURRENT FILING DATE: 2002-07-23
 ; NUMBER OF SEQ. ID NOS: 24
 ; SOFTWARE: Microsoft Office 97
 ; SEQ. ID NO. 23
 ; LENGTH: 2151
 ; TYPE: DNA
 ; ORGANISM: artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: mutant from Rhodotorula glutinis
 US-10-138-970A-23

Query Match 49.5%; Score 1196.6; DB 15; Length 2151;
 Best Local Similarity 74.0%; Pred. No. 2,4e-282;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

Db 99 CGCGGCTCGAGCAAGTGGGCTGCGGAGCGCGCACTTGGGCTTCTGCGCGCGAGCGCGCT 158
 Db 45 CGTGGATTCGCAAGAGAGCTGTCAATGCGCTTGAACCACTTCTGAGTGGAGGCTTC 104
 Qy 159 CGATGCGCAGCGCGCGCACAGTGCAGCTGAGATGTGACAGAGCTTCTTAAGGACCC 218
 Db 105 GCACTGCGCCACCAACCAAGTGCAGAGGTGACATGTGAGAAAGATGTCTGCGCGGCC 164
 Qy 219 CACCGAGAGCTGTGAGAGCTCAAGGAGTACAGCTCAACCGCTGAGAGCTTCTGCGGCG 278
 Db 165 GACCGAGTGAAGCTGAGAGCTGAGAGCTTCTGAGAGCTTCTGAGAGAGTGTCTGCGC 224

PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1998-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 9
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodotorula glutinis mutant
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2151)
OTHER INFORMATION:
US-10-188-523B-9

Query Match 49.5% Score 1196.6; DB 15; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2.4e-282;
Matches 1555; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 GCGCGCTCGACCAAGTCGGCTGCGGGCCCACTTCGGCTTCGCCCGCAAGCCCGGCT 158
DB 45 CGTCGATCCGCAAGCAGGCTGTCAATGCGCCCTCGACCACTCGCAGTCGAGGCTC 104
QY 155 CGATGGCCACGCGCGGCACTCGCAGCTGGAATGTCGAGAGCTCTTCAAGCC 218
DB 105 GCACTGCCCAACACCAGTCAAGCAGTCACTGTCAGAAAGTGTCCCGCGCC 164
QY 219 CACCGACGACGTCTCGAGCTCAGCGGGTACAGCCTCACTCCGTGACGTTGTCGAGC 278
DB 165 GACCGACTCGAGCTCGAAGTCAAGCGGTACTGCTCACTCGAAGCTGCTTCGCG 224
QY 279 GCGCGCAAGGCGCAGGTCGCGGTCCGATCAGAACGACGAGATCCGCGACGCTCGA 338
DB 225 CCGCGAAAGGCGCAGGCTGTCGCGTCAAGACGACGAGATCGCTCAAGATTGA 284
QY 339 CAAGACGTCGACTTCCTCAAGGCGCGCTTCAGAACCTCGGTCAGGAGTCAAGCGG 398
DB 285 CAATCGGTGAGTTCTTGCGCTCGCACTCTCCATAGCGTCTAGGCGTCAAGCTGG 344
QY 399 TTTCGGTCTCGGCGCACAGAGACTGAGATGACAGCTCAGCTTCAGAAAGCGCTAT 458
DB 345 ATTGGCGGATCCGAGACACCCGACCGAGAGAGCCATCTGCTCAGAAAGCTCTCT 404
QY 459 CGAGACGACGCTGCGCGGTCGAGCGGAGCGTCCGTCCTTCAGCGTCCGAGCGG 518
DB 405 CGAGACGACGCTGCGGTCGAGCGGAGCGTCCGTCCTTCAGCGTCCGAGCGG 464
QY 519 CCTCGAGAACAGCTTCGCTCGAGTCTCGAGTCTCGCGGCGCATGATGATCGCTCAATC 578
DB 465 TCTCGAGAACTGCTTCCTCGAGTCTCGAGTCTCGCGGCGCATGATGATGATGATG 524
QY 579 GCTACCGGTGCGCATCTGCGCGGTCGCGTCTGCTGCTGAGGCGCTCAACAACTTCT 638
DB 525 CTGACCGGCGGCGCATCTGCGGTCGCGTCTGCTGCTGAGGCGCTCAACAACTTCT 584
QY 639 GAACACCGCATCAAGCCCATCTGCTCGGCTCGGCGGCTCATCTCGGCGGCGGCGCT 698
DB 585 CAACACCGCATCAAGCCCATCTGCTCGGCTCGGCGGCTCATCTCGGCGGCGGCGCT 644
QY 699 CAGCCGCTCTGTCATCTGCGCGGCGCATCAAGCGTCAAGCGTCAAGTTTCACT 758
DB 645 CTCTCTCTCTCTCACTGTCAGGCGGCTCAAGCGTCAAGCGTCAAGTTTCACT 704
QY 759 TTGACAGAGGAAACGAGAAATCATGTTGCGCGGAGGCAATCGCTGTTGGTCT 818
DB 705 CGTCCAGAGGAGGAGAGAAATCTGTACGCCCGAGGCGCATGCGCTCTTCAACT 764
QY 819 CGAGCACTCTCTCTCGCGCGGAGAGAGGCTCTGCTGTCAGAGGAGAGCGCGCTCTC 878
DB 765 CGAGCGCGCTCTCTCGCGCGGAGAGAGGCTCTGCTGTCAGAGGAGAGCGCGCTCTC 824
QY 879 GCGCTCATGCGGCGCGGAGAGGCTCTGAGAGCTGCGCATGCTCTCTCTCGAGGCG 938
DB 825 AGCATGAGGCGCGGCTCTGAGAGGCTCTGAGAGCTCTCTCTCTCTCTCTCTCTCTCT 884

QY 939 CTTGACGCTCTCAAGTGAAGGCGCATGCTGCGCGCAGAGGCTCTGCGCGCTCAT 998
DB 885 GCTACGCGCATGAGCGGTGAGAGGATGTGTGCGCGCGCGCTCTGTTCCACCCCTTCT 944
QY 999 CCAGAGCTGTCG 1058
DB 945 TCAGAGGTCAAGCG 1004
QY 1059 CTTTCGCGCTCTGCTGTTTCCGTTGAGACGAGAGAGAGTCAAGTCAAGAGAGCA 1118
DB 1005 CTTGAGAGGAGAGCGCGCTTCTGTTGACCATGAGAGAGAGTCAAGTCAAGAGAGCA 1064
QY 1119 GAGCATCTTTCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1178
DB 1065 GCGCATCTCTGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1124
QY 1179 GGTGAGAGACATGATGACAGCGCTTCACTGACTCTGCTCGAAGAACAA---CGAGCAG 1235
DB 1125 CGTCAGGAGACCTCATTCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1184
QY 1236 CGAACACCGCTCTGACGTCGAGAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1295
DB 1185 CGAACACCTCTCATGACGTCGAGAACAGACTTTCGACACAGCGCGCGCATTTTCAGGC 1244
QY 1296 GTGCGCTGTCGATTCATGATGAGAAACAGAGCTCGGACTCGCGCTCATGCGAGAGT 1355
DB 1245 TGCGCGTGTGCGCAACCATGAGAAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1304
QY 1356 CAATTTCAGCATGACCGAGTTCCTCAACGCTGCAATGAAACCGCGCGCTGCTTCTG 1415
DB 1305 CAATTTCAGCATGACCGAGTTCCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 1364
QY 1416 CTTGCGCGAGAGACCGCGCTGCTCAACTATCAAGCGCGCGCGCGCGCGCGCGCGCGCG 1475
DB 1365 CTTGCGCGCGAGAGACCGCGCTGCTGCTCTCTCACTGAGAGGCTCTGACATGCGCG 1424
QY 1476 TGCTTACGCTTGAAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1535
DB 1425 GAGCTTACCTCGAGAGTGGAGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1484
QY 1536 AGAGATGGAGAACAGAGCGGTCACTGCTGCTCATCTGCGCGCGCGCGCGCGCGCG 1595
DB 1485 TGAATGCGAGACAGAGCGGTCACTGCTGCTCATCTGCGCGCGCGCGCGCGCGCG 1544
QY 1596 GCGCAAGAGCTCTTCT 1655
DB 1545 GTCGAAGAGCTCTTCT 1604
QY 1656 GAGCTCTGCGCGAGTGAAGTCTGACTTCAAGAGAGTTCGACCGGCTTCTCCGACTCT 1715
DB 1605 GAGCTTCTGCGCGAGTGAAGTCTGACTTCAAGAGAGTTCGACCGGCTTCTCCGACTCT 1664
QY 1716 CTTTCAGAGACCTCTGAGACTGAGCTCGACCTCA-----ACGCACTTTCGCT 1763
DB 1665 CATGACAGACCTTTCGCTCGCGCATGACCGGCTGAACTCGCGCGAGAGTGTGGA 1724
QY 1764 CGAGTCAAGAGGCGCTCAACAGGCTTTCGAGCAGAGAGAGAGTTCGAGCTTGAAGC 1823
DB 1725 GAAAGTGAAGAAAGCTTCGCGCAAGGCGCTTCAGAGAGAGCAACTGTCAGCTTCGCT 1784
QY 1824 GCGCTGAGAGAGAGCTTCTCGTACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1883
DB 1785 GCGTGGCAGAGAGCTTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1844
QY 1884 GCGCTTTCGCAAGCTCACTTTCGCTGACAGCGGTGAAGGTTGCTTGGCGAGAA 1943
DB 1845 GTC-----GCTCTGCTGCGCGCGCGCTCAACGCTGGAAGGTGCGCGCGCGCGCT 1895
QY 1944 GCGCATCTGCTGACGCGGAGGTGAGAGCGCTTTCGAGAGAGCGCGCTTTCGAGAGC 2003
DB 1896 GCGCATCTGCTGACCGCGCAAGTTCGAGAGCTTTCGAGAGCTTTCGAGAGCTTTCGAGAGC 1955

2004 GCCGGCGGACGATCTCTTCGCGCGGACAGCGCGTCTGTACTGTTGTGGGAGGA 2063
1956 GCCGCGCTCTCGTACTCTTCGCGCGGCACTAGATCTCTAGCTTGTGTCGGAGGA 2015
2064 GCTCGGCGTGGAGGCGCGCGCGGACGCTGTTGTGCGGTCGACAGAGACATCGG 2123
2016 GCTTGGGTCAGAGCCCGCGCGGAGACGTTCTCTTCGAGAGAGAGTACATCGG 2075
2124 GAGCAACGTCCTCGGCACTTACAGAGGCAATCAAGAGCGCGCATCAACCATGCTCGT 2183
2076 CTCGACGTCCTCCAGATCTTACGAGGCGCATCAAGTGGGAGAGATCAAGATGCTCTCT 2135
2184 CAAGATGCTGCGCTA 2198
2136 CAAGATGCTGCTTA 2150

RESULT 8
JS-10-374-366-3
Sequence 3, Application US/10374366
Publication No. US20040014085A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: C11794 US NA
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/360,279
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO: 3
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodospiridium glutinis
JS-10-374-366-3

Query Match 49.5%; Score 1196.6; DB 16; Length 2151;
Best Local Similarity 74.0%; Pred. No. 2,4e-282;
Matches 1665; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

2Y 99 CGCGGCTCCGACCAAGTGGAGTGGGGGCCCACTTGAGCTTCTCGCGGACGCCCGGCT 158
45 CGTGCATCCGGAAGGACGAGCTGCAATGGCGCTCGACCAACTCGAGTGGCAGGCTT 104
2Y 139 CGATGGCCACCGCGGACCAAGTGGAGTGGAGAGTGGTCTCGAGAGCC 218
105 GCACCTGCCCAACCCAGGTCACGCAAGTGCATGTCGAGAGATGCTGGCGGCT 164
219 CACGACACGTCGTGAGCTCAGCGGCTACAGCTCACGCTCCGTGAGTGGCGCG 278
165 GACCGATCGAGCTCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 224
279 CGCGCGCAAGGGGCGCAAGGTCGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 338
225 CGCGAGAGGGGCGCAAGGTCGCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284
339 CAAGAGCGTGCATCTTCCTCAAGAGCCAGCTTCAAGATCGGTCTACAGAGTCAAGAG 398
285 CAATATGTCGAGTCTTCGCGTCCAGATCTTCATAGAGGTCACAGGTCAGAGTGG 344
399 TTTGCGTGGCTGGCGGACACAGAGTGAAGATGCACTGAGCTCCAGAGAGGCTCAT 458
345 ATTGGCGGATCCGAGACACCCGACAGAGAGCGCATCTCGCTCCAGAGAGGCTTCTCT 404
459 CGAGCACGAGCTTGGCGGTCAGCGCGAGCGCGCTCGCTTCGAGCGTGGAGCGGG 518
405 CGAGCACGAGCTTGGCGGTCAGCGCGAGCGCGCTCGCTTCGAGCGTGGAGCGGG 464
519 CGTGGAGACAGCTTGGCGGTCAGCGCGAGCGCGCTCGCTTCGAGCGTGGAGCGGG 578
465 TCTCGAGAACTGCTTCCCTCGAGGTTGTGGCGGCGCATGACATCGCGTCAAGAG 524

QY 579 GCTCAGCGGAGCACTGGGCGGTCGCGCTGATGCTTGAAGGAGCTTCAACCACTTCT 638
DB 525 CTGACCGGAGGCACTTGGCTGTGCGCTCTGCTGCTCGAGGCGGCTTCAACCACTTCT 584
QY 639 GAACCAACGATACAGCCCATGCTTCCCTCCGCGCTTCAATCTGGGCTGGGAGACT 698
DB 585 CAACCAAGGATACAGCCCATGCTTCCCTCCGCGGCACTTCTGCGTGGGAGACT 644
QY 699 CAGCGGCTTCTGTAATGCGCGCGGATCAACCGGTCAACCCGAGTCAAGTTCAGT 758
DB 645 CTCTCTCTCTTCAATGCAAGGCGCATGAGGTCACCCGAGAGAGTTCAGT 704
QY 759 TTTGACAGAGAGCAAGAGATCATGTTTGGCGGAGGCAATCTCGCTTTGTGT 818
DB 705 GGTTCAGAGAGGCAAGAGATCTGTACGCGCGGAGGAGTGGCTTCAACCT 764
QY 819 CAGAGCATGCTCTTGGCGCGGAGAGGTCGCTGCTGCTCAACGAGGCGGCTTC 878
DB 765 CAGAGCGTGTCTTGGCGCGGAGAGGTCGCTGCTGCTCAACGAGGCGGCTTC 824
QY 879 CAGCTGATGAGCAACCTTCACTGACGATGCAATGCTCTGCTCTCTGAGG 938
DB 825 AGCATGATGAGCAACCTTCACTGACGATGCAATGCTCTGCTCTCTGAGG 884
QY 939 CTGAGCGCTTCACTGAGGAGGATGCTGCGGCGAGGCTGCTTGGCGGCTTCA 998
DB 885 GCTCAGCGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAG 944
QY 999 CAGAGAGTGTGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1058
DB 945 TCAGAGAGTCAAGGCGGCTTCAACGAGGATGAGGATGAGGATGAGGATGAGGAT 1004
QY 1059 CCTTCCGCTGCTGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1118
DB 1005 CTTGAGAGAGAGGAGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1064
QY 1119 GAGGATTTTGGCGGAGAGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1178
DB 1065 GGGGATTTTGGCGGAGAGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1124
QY 1179 GGTGAGAGATGATGAGGAGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1235
DB 1125 GGTGAGAGATGATGAGGAGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1184
QY 1236 GCAACACCGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1295
DB 1185 GCAACACCGGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTC 1244
QY 1296 GTCGCTGTCTCATTTTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1355
DB 1245 TGGCGTGTGCGCAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
QY 1356 CAATTCAG 1415
DB 1305 CAATTCAG 1364
QY 1416 CTTGCTGCGGAG 1475
DB 1365 CTTGCTGCGGAG 1424
QY 1476 TGTTCAGCTTTCGAGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1535
DB 1425 GGTGATGAG 1484
QY 1536 AGAGATGAG 1595
DB 1485 TGAATGAG 1544
QY 1596 GGTGATGAG 1655
DB 1545 GTTCAAG 1604

QY 1656 CGACCTCCGCGCATGAGAGCTGACTTCAAGAGAGTTGACCCGCTTCCCGGACTCT 1715
 Db 1605 CGACTTGGCGGCGACCGAGTTGAGTTCAAGAGAGTTTGGCCCGGACCATGCTGCT 1664
 QY 1716 CTTCCAGACGACCTCGGCACTGGCTCGACGTCA-----ACGCACTTGGGCT 1763
 Db 1665 CATCGACGACCTTGGCTCGGCACTGACCGGCTGAACTGGCGGACGAGCTGCGA 1724
 QY 1764 CGAGGTCAAGAGGCGCTCAAGAGTTTCCAGAGAGAGAGTTACGCTTCCGAGCC 1823
 Db 1725 GAAGGTGAACAGAGCTTCCGAGAGGCTTCAAGAGAGCAACTGTAAGACTTGTGCC 1784
 QY 1824 GCGCTGGACAGAGCGCTTCTCGTACGAGAGCGGACCGCTCGAGCTCCTGCTCTC 1883
 Db 1785 GCGCTGGACAGAGCGCTTCTCTTCCGCGCGGACCGCTCGAGGCTCTGCTGTGAC 1844
 QY 1884 GCGCTGGACAGAGCTTCTCTTCCGCGCGGACCGCTCGAGGCTCTGCTGTGAC 1943
 Db 1845 GTC-----GCTCTGCTGCGCGGCTCAAGAGTGGCGCGCGGAGTC 1895
 QY 1944 GCGCATCTGCTCAAGCGCGGAGTGGCGAGCGCTTCCGAGAGAGCGCGCTTCCGAGG 2003
 Db 1896 GCGCATCTGCTCAAGCGCGGAGTGGCGAGCGCTTCCGAGAGAGCGCGCTTCCGAGG 1955
 QY 2004 GCGCGGACGAGATCTCTGCGCGGAGCGCGCTCTCTTCTGCTTGGCGGAGGA 2063
 Db 1956 GCGCGGACGAGATCTCTGCGCGGAGCGCGCTCTCTTCTGCTTGGCGGAGGA 2015
 QY 2064 GCTCGGCGTGGAGGCGCGCGGAGGAGTGGTGGCGGAGAGAGAGATCGG 2123
 Db 2016 GCTGGCGTGGAGGCGCGCGGAGGAGTGGTGGCGGAGAGAGAGATCGG 2075
 QY 2124 GAGCAAGCTCTGCGGATCTCAAGGCGCATCAAGAGCGAGCGGATCAAGCGCTCTG 2183
 Db 2076 CTGCAAGCTCTCAAGATCTCAAGGCGCATCAAGTGGGAGAGATCAAGCGCTCTG 2135
 QY 2184 CAAGATGCTCGGCTA 2198
 Db 2136 CAAGATGCTCGGCTA 2150

RESULT 9

US-09-939-408A-18
 ; Sequence 18, Application US/09939408A
 ; Patent No. US20020102712A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yoshida, Roberta
 ; APPLICANT: Koelstra, Anna
 ; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 ; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 ; FILE OF INVENTION: Using Same
 ; FILE REFERENCE: 29479/500NSCA
 ; CURRENT APPLICATION NUMBER: US/09/939,408A
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 09/624,693
 ; PRIOR FILING DATE: 2000-07-24
 ; PRIOR APPLICATION NUMBER: PCT/US01/23270
 ; PRIOR FILING DATE: 2001-07-24
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: Patent Ver. 2.0
 ; SEQ ID NO 18
 ; LENGTH: 2439
 ; TYPE: DNA
 ; ORGANISM: Rhodotorula toruloides
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2148)
 ; US-09-939-408A-18

Query Match 49.5%; Score 1196.6; DB 9; Length 2439;
 Best Local Similarity: 74.0%; Pred. No. 2,4e-282;
 Matches 1565; Conservative 1; Mismatches 525; Indels 24; Gaps 3;

QY 99 CCGCGCTCCGACCAAGTGGCTGGGCGCCCACTTGGCTCTCTCCGCGGACGCCGCT 158
 Db 45 CCGCGCATCCGCAAGGAGCGGCTGTCAATGGGCGCTCGACCAACTCGAGTCCAGGCTC 104
 QY 159 CGATGGACGCGCGGACGAGTGGAGCTCGAGATGTGTGAGAGAGTCTCTGAGACCC 218
 Db 105 GCACTGCGCCCAACCCAGGTCAGGAGTGTGAGATGTGTGAGAGAGTGTGTCCGCGCT 164
 QY 219 CACTGACGAGTGTGAGTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 278
 Db 165 GACGAGTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCT 224
 QY 279 CCGCGCGAGGCGGAGGAGTGGCGGCTCGAGAGAGAGAGAGATCCGCGGAGCTTCA 338
 Db 225 GCGAGAGAGGAGGAGGCTGTGGCTCAAGAGAGAGAGAGATCCGCTCAAGAGATTA 284
 QY 339 CAAGAGCTGAGTCTCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAGGCTCAAG 398
 Db 285 CAATGCGTCAAGTCTTGGGCTCGCAACTCCATGAGGCTTCAAGGCTCAAGGCTCAAG 344
 QY 399 TTTGGTGGCTGGCGGACAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 458
 Db 345 ATTGGGAGATCGGAG 404
 QY 459 GAGAGAGAGTGGCGGAG 518
 Db 405 GAGAGAGAGTGGCGGAG 464
 QY 519 CTTGAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
 Db 465 TCTGAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 524
 QY 579 GCTGAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
 Db 525 CTTGAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 584
 QY 639 GAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 698
 Db 585 CAACGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644
 QY 699 GAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 758
 Db 645 CTTGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704
 QY 759 TTTGAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816
 Db 705 CTTGAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764
 QY 819 GAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
 Db 765 GAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 824
 QY 879 GAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
 Db 825 GAGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 884
 QY 939 CTTGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
 Db 885 GCTCAAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
 QY 999 CCAAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1058
 Db 945 TCAAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
 QY 1059 CTTTCCGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
 Db 1005 CTTGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
 QY 1119 GAGAGAGTGGCGGAG 1178
 Db 1065 GAGAGAGTGGCGGAG 1124
 QY 1179 CTTGAGAGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1235

1125 CGTGAAGACCTATTACGCCCAACCGCCCTCTCACTACGAGCGCGCAGTCCAGCAC 1184
1236 CGACACCCCGCTCTCGACGTGAAGAACAGACCGCGCAAGGCGCACTTCCAGGC 1295
1185 CGAACCCCTCTCACTGACGATGAGAACAGACTTGGACACAGGCGGCAATTTCCAGGC 1244
1296 GTGGCTGTCTCGATTTCATGAGAGAACCAAGCTTCGACCTGCCTCATGCGCAAGCT 1355
1245 TGCCGCTGTGCGCAACCACTGAGAGAACTCCGCTCGGCTGCCAATGGCAAGT 1304
1356 CAACCTACGAGTGAACCGAGTTGCTCAACGCTGCCATGAACCGCGCTTCTTGCT 1415
1305 CAACCTACGAGTGAACCGAGTTGCTCAACGCGCGCATGAACCGCGCTTCTTGCT 1364
1416 CCGTGGCGGAGGACCGGTGCGTCAACTACGCGCAAGGCTTGAATTCATGACGCG 1475
1365 CCGTGGCGGAGGACCGGTGCGTCTCTTACCTGCAAGGCTTGAATTCATGACGCG 1424
1476 TGCTTACGCTTCGAGAGTGGCGCACTTTCGCAACCGCGTCACTTCTGTCAGCGCG 1535
1425 GCGCTACACCTCGGAGTTGGAGACCTCGCAACCGCTGACAGCAATGTCAGCGCG 1484
1536 AGAGATGGGACCAACGAGCGCTCAACTGCTGCTCTGATCTCGCGCGCGCGCACTGCG 1595
1485 TGAGATGGGACCAACGAGCGCTCAACTGCTGCTCTGATCTCGCGCGCGCGCACTGCG 1544
1596 GCGCAAGAGAGCTCTTCT 1655
1545 GTCCACACAGCTCTTCT 1604
1656 CGACCTCCGCGCAGTGAAGCTCACTTCAAGAGAGCTTCAAGAGAGCTTCAAGAGAGCTT 1715
1605 CGACTTGGCGCGATGAGTTCAGATTCAAGAGAGCTTCAAGAGAGCTTCAAGAGAGCTT 1664
1716 CCGTACAGAGAGCT 1763
1665 CATGACAGAGAGCTTCT 1724
1764 CGAGGTCAAGAGAGGCTCAACAGAGCTTCTGAGAGAGCTTCTGAGAGAGCTTCTGAGAG 1823
1725 GAGGTGAGAGAGGCT 1784
1824 GCGGTGAGAGAGGCT 1883
1785 GCGGTGAGAGAGGCT 1844
1884 GCGGTGAGAGAGGCT 1943
1845 GTC-----GCT 1895
1944 GCGCATCT 2003
1896 GCGCATCT 1955
2004 GCGGTGAGAGAGGCT 2063
1956 GCGGTGAGAGAGGCT 2015
2064 GCGGTGAGAGAGGCT 2123
2016 GCGGTGAGAGAGGCT 2075
2124 GAGCAACGCT 2183
2076 CTCGAAACGCT 2135
2184 CAAGATGCT 2198
2136 CAAGATGCT 2150

RESULT 10

US-09-765-873A-7
; Sequence 7, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: B01009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-7

Query Match 49.4%; Score 1195; DB 9; Length 2151;
Best Local Similarity 73.9%; Pred. No. 5.9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

QY 99 CGCGCTCCGACCAAGTGGGCTGGGCGCCGCACTTGGCTTCCGCGCAGCGCGCGCT 158
DB 45 CGTGCATCCGCAAGAGAGGCTGTCAATGCGCGCTCGACCACTTCGAGTGGCAGCTTC 104
QY 159 CGATGGCCAGCGCGCGCAGCAGTCCGAGTCCGAGTCCGAGGAGCTCTTCAGCAGCC 218
DB 105 GCACCTGCCCAACCAACGAGTCAAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 164
QY 219 CACCGACAGCTCTGAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 278
DB 165 GACCGATCCAGCGTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 224
QY 279 CGCGCGCAAGGCGCGCAGGCTCGCGCTCCGAAACGACGAGAGATCCGCGCAGCGCTCA 338
DB 225 CGCGAGAGGCGCGCGCTGCTCGCGTCAAGAGCAGGAGAGATCCGCTCAAGAGATGA 284
QY 339 CAAGAGCTGAGCTTCCGCAAGGCGCGCTTCAAGCTGAGTCAAGGAGTCAAGGAGTCA 398
DB 285 CAATCGGTGAGTCTTGGCGCTGCACTTCCATGAGGCTTCAAGGCTTCAAGGCTTCAAG 344
QY 399 TTTCGATGCTCGCGCAGCAAGAGACTGAGATGAGTCAAGCTTCAAGGCTTCAAGGCTTCA 458
DB 345 ATTGGCGGATCCGCAACCGCGCACCGAGAGCGCATCTCGCTCCAGAGGCTTCTCT 404
QY 459 CGAGCACAAGCTTGGCGCTGAGAGCGCGAGCTCCGCTCTCTTCAAGGCTTCAAGGCTTCA 518
DB 405 CGAGCACAAGCTTGGCGCTGAGAGCGCGAGCTCCGCTCTCTTCAAGGCTTCAAGGCTTCA 464
QY 519 CCTGAGAAACAGCTTCGCTCGAGTCTGTCGCGCGCGCGCATTCGCTCAAGCTTCAAGCTC 578
DB 465 TCTGAGAACTCGCTTCCCTCGAGTGTTCGCGCGCGCATTCGCTCAAGCTTCAAGCTC 524
QY 579 GCTCAAGGTGGCGCATTCGCGCGCTGCTGCTCTTGAAGGCTTCAAGCTTCAAGCTTCA 638
DB 525 CTGACCGCGCGCATTCGCGCTGCTGCTCTGCTGCTTCAAGGCTTCAAGCTTCAAGCTTCA 584
QY 639 GAAACACCGCATTCAGCGCGATTCGCGCTGCGCGCGCGCATTCGCTCAAGCTTCAAGCTTCA 698
DB 585 CAACACCGCATTCAGCGCGATTCGCGCTGCGCGCGCGCATTCGCTCAAGCTTCAAGCTTCA 644
QY 699 CAGCGCTCTCTGATCAAGTCCGCGCGCATTCAGCGCGCATTCAGCGCGCATTCAGCGCTCAAGCTTCAAGCTTCA 758
DB 645 GCTCTCTCTCTCAATTCAGCGCGCATTCAGCGCGCATTCAGCGCGCATTCAGCGCTCAAGCTTCAAGCTTCA 704
QY 759 TTTCGACGAGGAGACCGAGAGATTCGCTTGGCGCGAGGCGCATTCGCTTTCGCTTTCGCT 818

339 CAAGAGCTGACTTCTCTCAAGGCCAGCTTCAGAACTCGGTCTACGAGTCAACGAG 398
285 CAATTCGGTGAATTTCTGGGCTGCAACTCTCCATGAGGCTTACGGGGTCAACGACTG 344
399 TTTGGTGGCTCGGCGACACGAGAGTTCAGATTCAGCTTCAAGAGCGCTCAT 458
345 ATTGGGGGATCCGCAACACCGGACCGGACCGCATTCCTCTCAAGAGCTTCTCT 404
459 CGAGACACAGCTGCGGCGGTGAGCGCGACGCTCGCTCTTCTAGCGTGGAGCGCG 518
405 CGAGACACAGCTTCTCGGCTGTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 464
519 CTTGAGAAACGCTTCGCTCGAGGTGCTCGGCGCGCATGCTATTCGGCTCACTC 578
465 TCTCGAGAACTCGCTTCCCTCGAGGTTGTTGCGGCGCATGACATCCGCTCAACAG 524
579 GCTACCGGTGGGCACTCGGCGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
525 CTTGACCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 584
639 GAACACACGCTACAGCGCATCGTCCCTCGGCGCTCATCTCGGCTCGGCGAGCT 698
585 CAACACGAGTCAACCCCATCGTCCCTCGGCGCATCTGCGTGGGCGAGCT 644
699 CAGCCGCTCTCTGATTCGCTCGGCGCATCAACGCTCAACGCTCAAGTTTACGT 758
645 GTCTCTCTCTCTCAATTCGACGCGCATGAGGCTCAACGAGCAAGTTGACGT 704
759 TTTGACAGAGGAAACGAGAAATCATGTTGCGCGGAGGCTCATCGCTTTTGGTCT 818
705 GCTCCACAGAGGCAAGAGAGATCTGTACGCTCGGAGGCGATGCGCTTCAACCT 764
819 CGAGCAGCTGCTCGCTCGGCGCGGAGAGGCTCTGCTGCTCAACGAGCGGCTCTC 878
765 CGAGCGGCTGCTCTCGGCGCGGAGAGGCTCTGCTGCTCAACGAGCGGCTCTCTC 824
879 CGCTCGAGTGGCGACCTCATGCTGCAAGCTGCAATGCTCTGCTCTCGAGGCG 938
825 AGCATGATGCTCAACCTGCTGCTGCAAGCTGCAATGCTCTGCTCTCTCGAGT 884
939 CTTGACGCTCTCAACGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 998
885 GCTCAGGCTCATGAGGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
999 CCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1058
945 TCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1004
1059 CTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118
1005 CTTGAGGAGAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1064
1119 GGGCATTCTTCCGAGAGACGCTACCGGCTCGGAGCTGCTGCTGCTGCTGCTGCT 1178
1065 GGGCATTCTTCCGAGAGACGCTACCGGCTCGGAGCTGCTGCTGCTGCTGCTGCT 1124
1179 CGTGAAGAGCATGATGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
1125 CGTCAAGCAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184
1236 CGACACACGCTCTCTGAGCTGAGAAACAGACAGCGCGACGCGCACTTTCAGAG 1295
1185 CGACACACCTCTCTCATGAGCTGAGAAACAGCTTTCAGACAGCGCGCACTTTCAGAG 1244
1236 GTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
1245 TCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
1356 CAACCTTCAAGAGTCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
1305 CAACCTTCAAGAGTCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
1416 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1475

1365 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1424
1476 TGGTTCAGCTTGGAGCTGGAGCCTTGGCAACCGGCTGCACTTCTGCTGCTGCTGCT 1535
1425 GGGCTACCTTGGAGTGGAGCCTTGGCAACCGGCTGCACTTCTGCTGCTGCTGCTGCT 1484
1536 AGAGATGGGCAACAGGCGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1595
1485 TGAATGGCAACAGGCGGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1544
1596 GGGCAAGAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
1545 GTCCAGAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
1656 CGACTTCGCGCATGAGCTGCACTTCAAGAGCAGTTGCAACCGCTTCTGCTGCTGCT 1715
1605 CGACTTGGCGCATGAGTGGAGTTCAAGAGAGCTTGGCGCGCATGCTGCTGCTGCTGCT 1664
1716 CTTGACAGAGCTTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1763
1665 CATGACACAGCATTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1724
1764 CGAGTCAAGAGGCGCTCAACAGGCTGCTGAGCAGACAGACAGCTAGACCTGAGGCT 1823
1725 GAGGTGAACAGAGCTGCTGCAAGGCTGCTGAGCAGACAGACAGCTAGACCTGAGGCT 1784
1824 GCGTGGACAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1883
1785 GCGTGGACAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1844
1884 GCGCTTGGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1943
1845 GTC-----GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1895
1944 GGCATCTGCTTCAAGCGGAGGCTGCAACCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 2003
1896 GGCATCTGCTTCAAGCGGAGGCTGCAACCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1955
2004 GCGGCGCAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2063
1956 GCGGCGCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2015
2064 GCTGCGGCTGAGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2123
2016 GCTTGGGCTTCAAGGCGCGCGGCGGAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2075
2124 GAGCAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2183
2076 CTCGAAGCTTCTCAAGATCTACGAGGCTCAAGTGGGAGGATCAACAGCTTCTGCT 2135
2184 CAGATGCTGCGCTA 2198
2136 CAGATGCTGCTTA 2150

RESULT 12
US-10-374-366-1
Sequence 1, Application US/10374366
Publication No. US2004014085A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: C11794 US NA
CURRENT APPLICATION NUMBER: US/10/374,366
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/360,279
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 2151

TYPE: DNA
ORGANISM: Rhodospiridium glutinis
US-10-374-366-1

Query Match 49.4%; Score 1195; DB 16; Length 2151;
Best Local Similarity 73.9%; Pred No. 5,9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

QY 99 CCGCGCTCCGACCAAGTCGGCTCGCGCCCACTTCGGCTTCGCGCGACGCCGCGCT 158
DB 45 CGTCGATCCGAAAGCAGGCTGTGATGCGCCCTCGACCAACTCGAATCGAAGCTC 104
QY 159 CGATGCGCCACGCGCGCGACCACTCGACCTGAGATGTCGAGAGAGTCTCTCAAGACC 218
DB 105 GCACCTGCCCAACACCGAGTCAAGAGTCACTGTCGAGAAAGTCTCGCGCGCC 164
QY 219 CACCGACGATCTGCTGAGCTGAGGGGTACAGCTCCGCTGACGTTGTCGGCGC 278
DB 165 GACCGATCGACGCTGAACTGACGGCTACTCGCTCAACTCGGAACTGTCGTCGCG 224
QY 279 CGCGCGCAAGGGCGCAGGGTCCGCTCAAGAACGACGAGATCCGCGCAAGCTCGA 338
DB 225 CGCGAGAGAGGCAAGCGCTGTCGCGGTCAAGACAGGACGAGATCCGCTCAAGATTGA 284
QY 339 CAAGAGCGTGCATCTCTCAAGGCCAGCTTCAGAACCTCGGTCAAGGATCCACCGG 398
DB 285 CAATTCGTCGAGTCTTTCGCTCGCACTCTCAATGAGGTCCTACGCGCTCAAGCTGG 344
QY 399 TTTGCTGCTCGCGCGACACGAGACTGAGATGACCTCAAGCTCCAGAGCGCTCAT 458
DB 345 ATTTGCGGATCCGACACCCGACCGAGACGCACTTCGCTCGAGAGGCTCTCTCT 404
QY 459 CGAGACCAAGCTTCGCGCGGTGAGCGCGACCTCCGCTCTTGAAGCTCGGACCGG 518
DB 405 CGAGACCAAGCTTCGCGGTGTCCTCTCTGTCGTCGTCGCTTCGCTCGCGCGG 464
QY 519 CCTCGAGAACAGCTTCGCTCGAGGTGTCGCGCGCGCATGATGTCGCTCGCTCACTC 578
DB 465 TCTCGAAGAACTGCTTCCTCGAGGTGTCGCGCGCATGACAAATCGCTCAACAG 524
QY 579 GGTCAAGCTGCGCTGCGCTGCGCTGCTGCTGCTGAGCGCTCAACACTTCTT 638
DB 525 CTGACCGCGGCGCACTGCTGCTGCTGCTGCTGCTGAGCGCTCAACACTTCTT 584
QY 639 GAACACCGCATCAAGCGCATCTGTCGCCCTCGCGCGCTCATCTCGCGCGCGACCT 698
DB 585 CAACACCGCATCAAGCGCATCTGTCGCCCTCGCGCGCATCTCTGCGTGGCGCATCT 644
QY 699 CAGCGCGCTCTGATCATCGCGCGCATCAAGCGCTCAAGCGCTCAAGGTTTCACT 758
DB 645 GTCTCTCTCTCTCAATTCGAGCGCGCATCAAGCGCTCAAGCGCTCAAGGTTTCACT 704
QY 759 TTTGACAGAGGAAACGAGAAATCATGTTGCGCGGAGGCGCATCTCGCTCTTGTGCT 818
DB 705 CGTCAACAGAGGAAAGAAATCTGTAAGCGCGGAGGCGCATGAGCTCTTCAACT 764
QY 819 CGAGCACTCTGCTCGCGCGCGAGAGGCTCTGCTGCTGATCAAGGAAAGCGCGCTCTC 878
DB 765 CGAGCGCTCTGCTCGCGCGCGAGAGGCTCTGCTGCTGATCAAGGAAAGCGCGCTCTC 824
QY 879 CGCTCATTCGCGCGCGCGCATCTGATCTGACACTGCTCTGCTCTCTGCGAGCG 938
DB 825 AGCATGATGAGCGCGCTCTGCTGACGAGCTCAATGCTCTCTCTCTCTGAGTGC 884
QY 939 CTGACAGCTCTCAAGGTGAGGCGATGTCGCGCGAGAGGCTCTGCTGCGCGCTTAT 998
DB 885 GCTCAAGCGCATGAGGCTGAGAGGATGTCGCGCGAGCGCGCTCTTCAACCTTCTT 944
QY 999 CCAAGAGCTCTGCT 1058
DB 945 TCACAGAGTCAAGCGCGCTCAAGCGCGAGTCAAGTCAAGCGCGCGCGCGCGCGAGCT 1004
QY 1059 CTTTCGCGCTGCTGCTTTCGCTGAGCAGAGAGGTCAGTCAAGTCAAGAGAGAGA 1118

DB 1005 CCTCAGAGGAAAGCGCTTCTGCTGTCACATGAGAGAGAGTCAAGGTCAAGAGACAGCA 1064
QY 1119 GGGCATTTCTTGCGAGAGCGCTACCCGCTCCGACAGTGGCGCTCAATTCCTCGCGCGCT 1178
DB 1065 GGGCATTTCTTGCGAGAGCGCTACCCGCTTTCGCGAGTCTCTCAATGAGTGGCGCGCT 1124
QY 1179 CGTGAAGACATGATGACCGCTACTGACTCTCTGCTGAGAAACAAG---CGACGAC 1235
DB 1125 CGTCAAGGACCTTATTAACCGCCACCGCGCTCTCAACATGAGAGCGCGCAATGACGAC 1184
QY 1236 CGACAAACCGCTCTCTCACTGTCGAGAAACAAGACACCGCGCAAGCTTCTTCAAGCG 1295
DB 1185 CGACAAACCTTCAATGACGCTGAGAAACAAGCTTCTGACACAGCGCGCAATTTCCAGCG 1244
QY 1296 GTGCGCTGCTCGATTTCATGAGAAAGACAGGCTCGCACTCGCGCTCATGCGCAAGCT 1355
DB 1245 TGCGGCTGTGCGCAACACATGAGAAAGCTCGCTTCGCGCTGCGCAATGCGCAAGCT 1304
QY 1356 CACTTCAAGCATGACACCGAGTTGCTCAAGCTGCTCAAGCGCGCTTCTGCTGCT 1415
DB 1305 CACTTCAAGCATGACACCGAGATGCTCAAGCGCGCAAGAACCGCGCTTCTGCTGCT 1364
QY 1416 CTTGCGCTGCGAGAGACCGCTCGCTCAACTCAAGCGCGCAAGGCTTGAATTCACATCGC 1475
DB 1385 CTTGCGCGCGAAGACCGCTCGCTCTCTCACTCACTCAAGGCTGTCGATGCGCGCTGC 1424
QY 1476 TGCTTACGCTTCGAGCTCGAGCTCGACCTTGCAACCGCGTCACTACTTCTGTCAGCGCG 1535
DB 1425 GCGGTACACTCGAGAGTGGGACACTCGCGCAACCTGATGAGAGCATGTCACCGCG 1484
QY 1536 AGAGATGGCAACCGAGCGCTCACTGCTGCTCTCACTCTCGCGCGCGCACTGCGCA 1595
DB 1485 TGAATGGCGCAACCGAGCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1544
QY 1596 GGGCAAGAGCTCTTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
DB 1545 GTCAAGAGAGCTCTTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1604
QY 1656 GGACTCTCGCGCGATGAGCTGACTTCAAGAAAGCATGTCAGCGCTTCTTCCGACTCT 1715
DB 1605 GGACTCTCGCGCGATGAGCTGACTTCAAGAAAGCATGTCAGCGCTTCTTCCGACTCT 1664
QY 1716 CTTGAGCAGACCTCGGCACTGCGCTCGACGCA-----AGCACTTGGCGCT 1763
DB 1665 CATGACACGACCTTGGCTCGGCTGACATACCGGCTGAACTGCGCGAGAGCTGCTGA 1724
QY 1764 GAGGTCAAGAGCGCTCAACAGGCTCTGAGAGACGAGACGATACGATTCGAGCGC 1823
DB 1725 GAGGTCAAGAGCGCTCAACAGGCTCTGAGAGACGAGACGATTCGAGCGCTGCTGCT 1784
QY 1824 GCGCTGAGAGAGCGCTTCTGAGCGAGCGCGACCGCTGCTGAGCTCTCTGCTGCT 1883
DB 1785 GCGTGGACAGAGCGCTTCTCTGCGCGCGCGACCGCTGCTGAGGCTCTCTGCTGAC 1844
QY 1884 GCGCTTGGCAAGCTCAACCTTACTGCGCTCAACCGGTGAGAGGTTGCTGCGCGAGAA 1943
DB 1845 GTC-----GCTTGTCTGCGCGCGCTCAACCGCTGAGAGGCTGCGCGCGAGATC 1895
QY 1944 GGCATCTGCTTACGCGCGGAGGTGCGCAACCGCTTCTGCGAGAGCGCTTCTGCGAGCG 2003
DB 1896 GGCATCTGCTTACCGCGCAAGTCCGAGAACCTTCTGCTGCTGCGCGCTGAGACTGCTC 1955
QY 2004 GCGCGCGAGCATACCTCTGCGCGGACGCGCTCTGTAATCTGCTGCTGCGCGAGAA 2063
DB 1956 GCGCGCGCTCTGTAATCTCTGCGCGGACGCTGATCTCTGAGCTTCTGCTGCGAGAA 2015
QY 2064 GCTGCGCTGACAGCGCGCTGCGCGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2123
DB 2016 GCTTGGGCTCAAGCGCGCGCGGAGAGGCTCTTCTTCTGCGCAAGGAGGAGATCGG 2075
QY 2124 GAGCAAGCTCTGCGCGATCTAGAGCGCATCAAGAGCGCGCGCATCAAGCGCTCTGCT 2183

Db 2076 CTGACAGCTCTCCAGAGTTTACAGAGGCATCAAGTCGGCGAGAGTACACAGCTCTCT 2135
Y 2184 CAAGATGCTCGCGTA 2198
Db 2136 CAAGATGCTCGCTTA 2150

RESULT 13
JS-10-374-366-9
Sequence 9, Application US/10374366
Publication No. US20040014085A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: C11794 US NA
CURRENT APPLICATION NUMBER: US/10/374,366
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/360,279
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 9
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodospiridium glutinis
JS-10-374-366-9

Query Match 49.4%; Score 1195; DB 16; Length 2151;
Best Local Similarity 73.9%; Pred. No. 5,9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

Y 99 CGCGCTTCGACCAAGTCGGCTCGGGCCCCACTTGGCTCTCCGCCGACCGCGGCT 158
Db 45 CGTGCATCCGCAAGAGGCTGTCAATGGCGCTCGCCACCACTCGAGTCGACGGCTC 104
Y 159 CGATGGCAGCGCGCGACGACGATGCGAGTGTGACAGAGCTCCTCGACGCC 218
Db 105 GCACCTGCCCAACACGAGTCAAGTCGACATGTGAGAAATGCTCGCGCGC 164
Y 219 CACCGACGACGCTCGTGAAGTCAAGCGGTACAGCTTCAAGCTTGTGCGCGC 278
Db 165 GACCGACGACGCTCGTGAAGTCAAGCGGTACAGCTTGTGCGCGC 224
Y 279 CGCGCGCAAGGGGCGAGGTCGCGCTTCGACAGAGTCAAGATCCGCGACGCTCA 338
Db 225 CGCAGAGAGGAGGCGGCTGTCCGCTCAAGAGACGAGATCCGCTCAAGATTA 284
Y 339 CAAAGCGCTGACTTCTCAAGGCGCACTTCAAGCTTCAAGCTTCAAGAGTCAACGCG 398
Db 285 CAAATCGGTGAGTTCTTGGCGCTCGCACTCTCAATAGAGTCTTCAAGGCTCAAGCTGG 344
Y 399 TTTGGTGGCTCGCGCGACACGAGATGAGATGAGTCAAGCTTCAAGAGGCTCAT 458
Db 345 ATTGGCGAGTCCGAGACACCGCACCGAGAGCGCATCTCGCTCAAGAGCTTCTCT 404
Y 459 CGAGGACAGTCTCGCGCGTGAAGCGGACGCTCGCTTCAAGCTTCAAGAGGCGCG 518
Db 405 CGAGGACAGTCTCGCGGTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 464
Y 519 CCTCGAGAACGCTTCGCTGAGAGTGTGTCGCGGCGCGCATGCTATCGCGTCAATC 578
Db 465 TCTCGAGAACGCTTCGCTGAGAGTGTGTCGCGGCGCGCATGCAATCGCGCTCAACAG 524
Y 579 GCTCAGCGGTGCTCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
Db 525 CTTGACCGCGCGCATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 584
Y 639 GAACACCGCATCAAGCGCATGCTCCGCTCGCGGCTCATCTCGCGTGGCGACCT 698
Db 585 CAACACCGCATCAAGCGCATGCTCCGCTCGCGGCGCATCTCTGCGTGGCGACCT 644
Y 699 CAGCCGCTCTGTAATGCGCGGCGCATCAAGCGTCAAGCGTCAAGGTTCAAGT 758

Db 645 CTCTCTCTCTCTCTATTCATTCAGCGGCATTCAGCGGTACCCCGACAGAGTCAAGT 704
Y 759 TTTGACAGAGGACCGAGAAATATGTTTGGCGCGAGGCGCATCTGCTTTGGTCT 818
Db 705 CTTCCACAGAGGACAGAAATATCTGTACCGCCGCGAGGCGATGCGCTTCAACT 764
Y 819 CGAGGAGTGTCTCGCGCGCGAGAGGAGTCTCGGTGTGTCAAGAGAGGCGGTCTC 878
Db 765 CGAGCGGTGTCTCGCGCGCGAGAGGAGTCTCGGTGTGTCAAGAGAGGCGGTCTC 824
Y 879 CGCGTGAATGGACCGCTGAGTGTGACGACTGCAATGCTCTGCTCTCTCGAGGC 938
Db 825 AGCATGATGACCGCTGCTGCTGCAAGCAGACATGCTCTGCTCTCTCGAGTC 884
Y 939 CTTGAGGCTCTTCAAGGTGAGGCGCATGCTGCGACGAGGCGTGTGCGCGGTAT 998
Db 885 GCTCAGGCGCATGAGTGTGAGGAGTGTGCGACGCGGCGGTGTGACCGCTTCTCT 944
Y 999 CCAAGAGTGTGCGCGCGCGACCGCGAGGTGAGGTGCGCGCAATCCGACGCT 1058
Db 945 TCAGAGCTACCGCGCTTCACCGACGAGATCAAGTCTCGGGAATATCGCAGCT 1004
Y 1059 CTTTCGAGTGTGCTTGTGCGTTGAGGACAGAGAGAGTCAAGTCAAGGACGACGA 1118
Db 1005 CTTGAGGAGAGCGGCTTGTGCTTCCACCATGAGAGAGAGTCAAGTCAAGGACGACGA 1064
Y 1119 GGGGATTTCTGCGCAAGACCGCTACCGCGCTCGAGAGTGTGCTGCTGCGCGCT 1178
Db 1065 GGGGATTTCTGCGCAAGACCGCTACCGCGCTCGAGAGTGTGCTGCTGCGCGCT 1124
Y 1179 CGTGAAGATGATGACAGCTACTGACTCTCTGCTCGAAGAACAA---CGAGAC 1235
Db 1125 CTTGAGGAGCTTATTCAGCGCGCGCGCTCTCAAGTCAAGGCGCGGCAATGACGAC 1184
Y 1236 CGACCAACCGCTCTCTGACGCTGAGAACAGACGCGCGCACTTCAGGCG 1295
Db 1185 CGACCAACCTCTATGACGCTGAGAACAGACTTGCACACGCGCGCAATTCAGGCG 1244
Y 1296 GTGCGCTGTCTGATTTGATGAGAGAACAGGCTGAGCTGCGCTCATGCGGACGCT 1355
Db 1245 TGCGGCTGTGCGCAACCAATGAGAGAGAGTGTGCGGCTGCGCAATGCGAGGCT 1304
Y 1356 CAACTTACGACAGTGCACAGAGTGTGCAACGCTGCAATGAACCGGAGCTGCTTGTG 1415
Db 1305 CAATTTACCGAGGTACAGAGTGTGCAACGCTGCAATGAACCGGAGCTGCTTGTG 1364
Y 1416 CTTGCTGCGAGAGACCGTGTCTCAATTCACGCGAGAGGCTTGAATTCATTCG 1475
Db 1365 CTTGCGCGCGAGAGACCGTGTCTCTCAACACTGCAAGGCGCTGAGATGCGGCTGT 1424
Y 1476 TGTTCAGCTTGGAGGCTGCGGCACTTGGCAACCGGTCACTTACCTTGTCCAGCGCG 1535
Db 1425 GCGCTTACCTTGGAGGTTGAGACACTTGGCAACCTGTGACAGCAAGTGTCCAGCGCG 1484
Y 1536 AGAGATGAGCAACAGGCGGTCAACTGCTGCTCAATCTCCGCGCGCGCATCGCGA 1595
Db 1485 TGAATGAGCAACAGGCGGTCAACTGCTGCTCAATCTCCGCGCGCGCATCGCGA 1544
Y 1596 GCGCAAGAGCTCTTCT 1655
Db 1545 GTCCAGAGAGCTCTTCT 1604
Y 1656 CGACCTCGCGCGATGAGGCTGCACTTCAAGAGAGGAGTTCAGACCGCTTCTCCGACTCT 1715
Db 1605 CGACTTGGCGCGACAGAGTTCAGATTCAAGAGAGGAGTTCAGACCGCTTCTCCGACTCT 1664
Y 1716 CTTGAGAGGACCTTGGGACGTGCGCTTCAGAGTCA-----AGGCAATTGCGCT 1763
Db 1665 CATGACAGGACCTTGGGACGTGCGCTTCAGAGGAGGAGTTCAGACCGCTTCTCCGACTCT 1724
Y 1764 CGAGGTCAAGAGGCGGTCAACAGGCTTTCAGAGAGAGGAGGAGTTCAGACCGCTTCCAGGCG 1823

Db 1725 GAAGGTGAACAAAGAGCTGCGCAAGCGCTCGAGCAGACCAACTCGTACGACTCGTCC 1784
QY 1824 GCGCTGACAGAGCGCTTCTCGTACGAGACGAGACCGTCTGAGCTCTCTGCTC 1883
Db 1785 GCGCTGACAGAGCGCTTCTCTCTCGCGCGGACCGTCTGAGCTCTCTCTGCTC 1844
QY 1884 GCGCTTGCACAGCTCAGCTTCTCTCTGAGCTCTCTCTGAGCTCTCTCTGAG 1943
Db 1845 GTC-----GCTCTGCTCTCGCGCGCTCAAGCGCTGAGAGTCCCGCGCGAGTC 1895
QY 1944 GGCATCT 2003
Db 1896 GGCATCT 1955
QY 2004 GCGCGGACAGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2063
Db 1956 GCGCGGCT 2015
QY 2064 GCTCGCGTGAAGCGCGCGCGCGCGAGTGTCTTGTGTGTGTGTGTGTGTGTGT 2123
Db 2016 GCTTGGCTCAAGCGCGCGCGCGAGAGCTTCTCTCTCTCTCTCTCTCTCTCT 2075
QY 2124 GAGCAAGCT 2183
Db 2076 CTGCAAGCT 2135
QY 2184 CAAGATGCT 2198
Db 2136 CAAGATGCT 2150

RESULT 14
US-10-374-366-13
; Sequence 13, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodospiridium rubrum
US-10-374-366-13

Query Match 49.4%; Score 1195; DB 16; Length 2151;
Best Local Similarity 73.9%; Pred. No. 5.9e-282;
Matches 1564; Conservative 1; Mismatches 526; Indels 24; Gaps 3;

Db 285 CAATTCGATGATTTCTTGGCTCTGCAACTCTCCATGAGGTCTTACGGGCTGACGACTG 344
QY 399 TTTGGTGGCTCTGGGCGGACAGAGAGCTAGAGTGTAGCTTACGCTTCAAGAGGCTCAT 458
Db 345 ATTGGGAGATTCGCAACACCGGACCGAGAGAGGCTATCTGCTCTCAAGAGGCTCTCT 404
QY 459 CGAGCAGGCTCTGCGGCTGACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
Db 405 CGAGCAGGCTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 464
QY 519 CTTGAGAAACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578
Db 465 TCTGAGAACTCTGCTTCTCTGAGGTTGTTCTGCGGCGGCTGAGCACTCTCTGAGAG 524
QY 579 GCTCAGCGCTGAGGCTCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 638
Db 525 CTTACCGGCGGCTCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGG 584
QY 639 GAACCAACGATACCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 698
Db 585 CAACCAACGATACCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 644
QY 699 CAGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 758
Db 645 CT 704
QY 759 TTTGACAGAGGAGACGAGAGATCATGTTTGGCGCGAGAGGCTCTGCTCTTGTGTCT 818
Db 705 GGTCAAGAGGAGGAGAGAGAGATCTTATACCGCGGAGAGAGAGAGAGAGAGAGAG 764
QY 819 GAGGAGAGTCTCTCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878
Db 825 AGCATGATGAGGAG 884
QY 879 CGCCTGATGAGGAG 938
Db 825 AGCATGATGAGGAG 884
QY 939 CTGAGAGGCTCTGAG 998
Db 885 GCTCAGGAG 944
QY 999 CCAAG 1058
Db 945 TCAAG 1004
QY 1059 CTTTTCGAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1118
Db 1005 CTTGAG 1064
QY 1119 GGGCATTTCTGCGAG 1178
Db 1065 GGGCATTTCTGCGAG 1124
QY 1179 CGTGAAG 1235
Db 1125 GGTGAAG 1184
QY 1236 CGAGCAACCGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295
Db 1185 CGAGCAACCGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
QY 1296 GTGAGGCTCTGAG 1355
Db 1245 TGCAGGCTCTGAG 1304
QY 1356 CAATTTCAAG 1415
Db 1305 CAATTTCAAG 1364
QY 1416 CTTGAG 1475

Db	1365	CCTCGGGGCCAAGACCCTCGCTCTCTACACCTGCAAGGGGCTCGACATCGCGCTGC	1474
2y	1476	TGCTTACGCTTGGAGCTCGGCACCTTGTCCAAACCCGGTCACTACCTTCTGTCAGCCGC	1535
Db	1425	GCGCGTACCTCGAGGTGGGACACCTTGGCAACCTGTGACAGGACATGTCAGCCGGC	1484
2y	1536	AGAGATGGGCAACACGAGCGCTCACTGCTCGCTCTCATCTCCGGCGCCGCCATGTCCGA	1595
Db	1485	TGAGATGGCGAACCAGGGGCTCACTGCTTGGCGTATCTCTGGGTCTGTCCGACGACCGA	1544
2y	1586	GGCCAAGCAGCTCTTCTTCTCTCTCTCTGCTCGCATCTGTACTGTACAGCTTCAGCCGT	1655
Db	1545	GTCCAAACACGCTCTTCTCTCTCTCTCGCACCCATCCCTCACTGCGCTTCTCCAAAGCAT	1604
2y	1656	CGACCTCCGCGGATGAGAGCTGCACTTCAAGAAGAGTTGCACCCCGGCTTCCCGACCT	1715
Db	1605	CGACTTGGCGCGACCGACCGATTCGATTTAAGAGAGATTTGGCCACGCATGCTCGCT	1664
2y	1716	CCTCCAGACGACCTCGGCACTGCGCTCGACGTCA-----ACGCATTTGCGCT	1763
Db	1665	CATCGACAGGACCTTTGGCTCGGCATGACGGGCTCGAAGCTGGCGGACGAGCTGTGCA	1724
2y	1764	CGAGTCAAGAAAGGGGCTCAATAAGCTCTGAGAGACGACGACGATGACGACTCGAGCC	1823
Db	1725	GAAAGTGAAACAAGAGCTCGCCAAAGCGCTCGAGACCAACTGTGACACTTGTGCC	1784
2y	1824	GCGCTGACGACGAGCGCTTCTCGTAGACGACGCGACCGCTCGTAGCTCTCTGCTCTC	1883
Db	1785	GCGCTGGACACGACGCTTCTCTTCTGCGCGCGGACACGCTCGTAGAGTCTCTGTGCAC	1844
2y	1884	GCCCTCTGCCAAAGTCACTTACTCTCGGTCAACGCTGAGAAAGTTGCTTGGCCCGAGAA	1943
Db	1845	GTCTCTGCTGCTGCGCGCGCTCAACCTGAGAAAGTGTGCGCGCGCGAGTCTC	1895
2y	1944	GCGCATCTCGCTCAGCGCGAGAGTGCAGAACCGCTTCTGAGAGAGCGCGTCTTGCAAGC	2003
Db	1896	GCGCATCTGCTCACCCTCGGACAGTCCGAGAGACTTCTGTGTCGCGCGGTGACCTGCTC	1955
2y	1956	GCGCGGCTCTGTAACCTCTCGCCGCGCACTAGATCTCTTAACGCTTCTGTCGCGAGGA	2015
Db	2064	GCTCGGCGTGCAGGGGCGCGCGGCGAGCGTGTTCGCGGTGCAGGAGAGATGCG	2123
2y	2016	GCTTGGCGTCAAGGCCCGCGGAGAGCTTCTCTCGGCAAGACAAAGGTGACGATCGG	2075
Db	2124	GAGCAACGTCTGCGCATCTACGAGGCGATCAAGAGACGCGCGCATCAACACGTCCTCGT	2183
2y	2076	CTCGAACGCTCTCAAGATCTTACGAGGCGATCAAGTCGGGCAAGATCAACACGTCCTCT	2135
Db	2184	CAAGATGCTCGGCTA 2198	
2y	2136	CAAGATGCTCGGCTTA 2150	
RESULT 15			
US-10-374-366-23			
Sequence 23, Application US/10374366			
Publication No. US20040014085A1			
GENERAL INFORMATION:			
APPLICANT: Tang, Xiao-Song			
APPLICANT: Milano, Joseph			
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS			
FILE REFERENCE: CL1/94 US NA			
CURRENT APPLICATION NUMBER: US/10/374,366			
CURRENT FILING DATE: 2003-02-26			
PRIOR APPLICATION NUMBER: 60/360,279			
PRIOR FILING DATE: 2002-02-26			
NUMBER OF SEQ ID NOS: 203			
SOFTWARE: PatentIn version 3.2.			
SEQ ID NO 23			
LENGTH: 2151			
TYPE: DNA			

! ORGANISM: *Rhodospiridium glutinis*
US-10-374-366-23

Query Match	49.4%;	Score 1195;	DB 16;	Length 2151;
Best Local Similarity	73.9%;	Pred. No. 5.9e-282;		
Matches 1564;	Conservative	1;	Mismatches 526;	Indels 24;
				Gaps 3;

QY	99	CGCGCTCCACCAAGTCGGGTGGGGCCCACTTGGCTCTCGCGGCACGGCCGCTC	158
Db	45	CGTGCATCCGCMAAGAGAGGTGTGCATGGGCGCTGCACCACTCGCAATCGAGCTC	104
QY	159	CGATGGCCACGCGCGGCACCAAGTCGACGCTTCGAGATGTGCAAGAGCTCTCAGCAACC	218
Db	105	GCACCTGCGCCACACCCAGGTCAAGCAGGTTCGACATGTGAGAAAGATGTCCGCGGCC	164
QY	219	CACCGAGAGCGTGTGAGCTCAAGCGGATCAGGCTCACCGTCCGATCAGTTGCGGCGC	278
Db	165	GACCGACTCGACGCTCGAATCGACGCGCTACTGCTTAACCTCGGAGACGTCGTTCCGC	224
QY	279	CGCCCGCAAGGCGCAGGTCGCGCTCCAAAGCAGCAGAGATCCGCGCACGCGTGA	338
Db	225	CGCAGAAAGGCGCAGGCGCTCTCGCGCTCAAGGACAGCGAGCAGATCCGCTCAAGATGA	284
QY	339	CAAGAGCTCGACTCTCTCAAGGCCAGCTTCAGAACTCGGTCTCAGGACATCCACGGG	398
Db	285	CAATCGGTCAAGTCTCGCGCTGCACTCTCCATGAGCGTCTCAGCGCTCACGACTGG	344
QY	399	TTTCGATGCTCGCGCCGACAAGAGCTGAGATGCAATGACCTCCAGAAAGCGCTCAT	458
Db	345	ATTGGCGGATCCGCAACACCGCGACAGAGAGCGCATCTCGCTCAGAAAGCTTCCT	404
QY	459	CGAGACCAAGCTTCGCGCGTGAACCGGAGTCCGCTCGTCTTCAGCGTCCGACCGG	518
Db	405	CGAGCACCAGCTCGCGGTGTCTCCCTTGGTCTTGAATCGTTCGCGCTCGCGCGCG	464
QY	519	CCTGAGAAACGCTTCGCGTCAAGGTCTCGCGGCGCATGATCATCGCGCTCACTC	578
Db	465	TCTGAGAACTCGCTTCCCTCGAGGTGTTCGGGGCCATGACAACTCGCGTCMAAG	524
QY	579	GCTACCGGTGGCACTCGGCGGTTCGCGCTTCCTCTTGAGGCGCTCACCACTTCTT	638
Db	525	CTTGACCGCGGCACATCGGTGTTCGCGCTGTGATCTCGAGGCGCTCACCACTTCT	584
QY	639	GAACCAACGCACTCAAGCCATCGTCCCCCTCCGCGGCTCATCTCGCGTGGAGGACCT	698
Db	585	CACCAAGGCACTCAACCCCATGTCCCTTCGCGGACCATCTCTGCTGGGGGACCT	644
QY	699	CAGCCGCTCTCTGATCATCGCGCGCGCATCACCGATCACCCCGACGTCAAGTTCACT	758
Db	645	CTCTCTCTCTCTAATGACAGCGGCATCAGCGGTACCCCGACAGAGGTGACGT	704
QY	759	TTTGACAGAGGAACCGAAGATCATGTTTGGCGGAGAGCCATCTCGCTTTGGCT	818
Db	705	CGTCACAGAGGCAAGAGAAATCTGTACGCCCGGAGGCGATGGCGCTTCAACT	764
QY	819	CGAGCAGTCTCTCGCGCGCAGAGAGAGGTCTCGCTTGATCAACGAAACGCGCTCTC	878
Db	765	CGAGCGCTCTCTCGCGCGGAGAGAGGTCTGCTCTCAACGGAACCGCGTCTC	824
QY	879	CGCTCGATGCGACCTCACTTGCACGATCGCAATGCTCTTCGCTCTCTTCGACGCG	938
Db	825	AGCATCGATGCGACCTCGCTCTGCACGACGACACATGCTCTCGCTCTCTCGACCTC	884
QY	939	CTTACCGGCTCTCAACGCTGAGGCGCATGTTCGCGAGAGGCTCGTTCGCGGCTTCAAT	998
Db	885	GCTACGGCGCATGACGTGAGAGGATGTGGCGCACGCGGCTCTTCACCCCTTCT	944
QY	999	CCACGACCTTCGCGCGCCGACCCCGGCGAGGTGAGGTGCGCGCAACATCGCACT	1058
Db	945	TCACGACCTCACGGCGCTCTCACCGACGACGATCGAATCGCGGAAACATCCCAAGCT	1004
QY	1059	CCTTTCGGCTGTGCTTTCGCTTGACAGAGAGAGGTCAAGTCAAGTCAAGGACGACGA	1118

Mon Sep 13 10:31:08 2004

us-09-939-408a-12.rnpb

Page 20

Db 1005 CCTCGAGGAGAACCGCTTTGCTGTCCACCATGAGAGAGGTCAAGGTCAAGGACGACGA 1064
QY 1119 GGGCAATTCTTCCGCGAGACCGGCTACCCGCTCCGACCGCTCCAGTTCTTCGGCCCGCT 1178
Db 1065 GGGCAATTCTCCGCGAGACCGGCTACCCGCTCCGACCGCTCCAGTTCTTCGGCCCGCT 1124
QY 1179 CGTGAAGACATGATGACAGGCTTACTCGACTCTCTCGCTCGAGAACACA---GAGGAC 1235
Db 1125 CGTCAAGCAGCTCATTCACGCGCCACGCGGTCTTCATCATGAGCGCGCCAGTGAAGAC 1184
QY 1236 CGACAAACCGGCTCTCTCGAGCTGAGAGAACAAAGCAGACCGCGACCGCGCACTTCCAGGC 1295
Db 1185 CGACAAACCGCTCTCATTCAGAGCTGAGAACAAAGCTTCGACCAAGCGCGCAATTCCAGGC 1244
QY 1296 GTGAGCTGTCTGATTTGATGAGAGAACAGGCTCGACCTGCGCTCATCGGCAAGCT 1355
Db 1245 TGCCTGTGTGGCCACACATGAGAGAACCTCGCTCGGCTCGCCAGATCGGCAAGCT 1304
QY 1356 CAATTTCACGAGTGCACCGAGTTGCTCAACGCTGCAATGAACCGCGGCTGCTTCGTG 1415
Db 1305 CAATTTCACGAGTGCACCGAGTTGCTCAACGCTGCAATGAACCGCGGCTGCTTCGTG 1364
QY 1416 CCTCGTGCAGAGACCGGCTCGCTCACTATCAAGGCAAGGCTTGGACATTCACATCGC 1475
Db 1365 CTTCCGCGCGAAGACCGCTCGCTCTCTTCACTGCAAGGCTTGCACATCGCGCTGC 1424
QY 1476 TGCCTTACGCTTGAAGCTGAGCCACCTTGCCACCGGTCATACCTTCGTCCAGCCGC 1535
Db 1425 GCGGTACACCTGAGGTTGGAGACCTTGCCACCTGTGACGACGATGTCCAGCGGC 1484
QY 1536 AAGATGGGGAACCGGCGCTCAACTCGCTGCTCATCTCCGCGCGCGGCACTGCCA 1595
Db 1485 TGAATGGGAACCGGCGCTCAACTCGCTGCTCATCTCCGCTGTGCAACGA 1544
QY 1596 GGGCAAGACGTCCTTCT 1655
Db 1545 GTCAACGACGTCCTTCT 1604
QY 1656 GCACTTCGCGGAGTGAAGCTGCACTTCAGAGACAGTTGACCCGCTTCCGACTCT 1715
Db 1605 CCACTTGGCGGAGCGGAGTTCAGTTCAAGAACAGTTGCGCCAGCCATCTCTCGCT 1664
QY 1716 CTTCCAGAGACCTCGGCACTGAGCTGACGTCA-----ACGCACTTGCCT 1763
Db 1665 CATTCACGACCTTTGCTCTCGCATGACCGGCTCGAACCTGCGGACGAGCTGTGCA 1724
QY 1764 CGAGTCAAGAAAGCGCTCAACAAAGCTCTTGAGACAGACGACGACCTCGAGCC 1823
Db 1725 GAAGGTGAACAAGCGCTCGCAAGCGCTCGACAGACCACTCGACACTCGTCC 1784
QY 1824 GCGCTGAGACAGCGCTTCTCTGTAAGCGGACCGGACCGGTGAGAGTCTCTCGCTC 1883
Db 1785 GCGTGGACAGAGCGCTTCTCTGCGCGCGGACCGGTGAGAGTCTCTCGTCA 1844
QY 1884 GCGCTTGCACAGCTCACTTACTGCGGTCAACGCGTGAAGGTTGCTCGGCGGAGAA 1943
Db 1845 GTC-----GCTCTGCTCGCGCGCGTCAAGCTGGAAGTGCAGCGCGCGGAGTC 1895
QY 1944 GGGCATCTGCTCAAGCGCGGAGTGAACCGCTTTCGACAGCGCTTTCGAGGC 2003
Db 1896 GGGCATCTGCTCAAGCGCGGAGTGAACCGCTTTCGAGTTCGCGCGGCTGACCTGTC 1955
QY 2004 GCGGCGACGACATCTCTGCGCGGACGCGGCTCTGACTCTGCTGCGGAGGA 2063
Db 1956 GCGGCGGCTCTGTAAGCTCTGCGCGGACCTCAAGATCTCTAGCCCTGTGCGGAGGA 2015
QY 2064 GCTGCGGCTGAGCGCGCGGCGGACGTTGTTGCGGCTGACAGAGACATCGG 2123
Db 2016 GCTTGGCTCAAGGCGCGCGGAGAGTCTTCTCGGCAAGAGGTGACGATCGG 2075
QY 2124 GAGCAAGCTTCGCGCATCTTACGAGGACATCAAGAGCGCGGATCAACAGCTCTGCT 2183
Db 2076 CTGCAAGCTTCGAAGTCTACGAGGACATCAAGTGGGAGATCAACAGCTCTCTCT 2135

QY 2184 CAAGATGCTGCGCTA 2198
Db 2136 CAAGATGCTGCTTA 2150

Search completed: September 12, 2004, 12:22:08
Job time : 1583.32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 11, 2004, 13:17:03 ; Search time 9123.4 Seconds
(without alignments)
7917.733 Million cell updates/sec

Title: US-09-939-408a-12

Sequence: 1 ctctgcctcactctaaccc.....gagttcaaaaaaaaaaaaaa 2419

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55028578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hcc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hcc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: em_gse_hum: *
18: em_gse_inv: *
19: em_gse_pin: *
20: em_gse_vrt: *
21: em_gse_fun: *
22: em_gse_mam: *
23: em_gse_mus: *
24: em_gse_pro: *
25: em_gse_rtd: *
26: em_gse_png: *
27: em_gse_vrl: *
28: gb_gse1: *
29: gb_gse2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	400.2	16.5	2598	11 AY103647	AY103647 Zea mays
2	386.2	16.0	2299	11 AY106831	AY106831 Zea mays
3	289.2	12.0	1864	11 AY104679	AY104679 Zea mays
4	268	11.1	859	14 CB632830	CB632830 OSJED11F

C 5	266.8	11.0	866	14 CB645202	CB645202 OSJNEB06M
C 6	266.6	11.0	860	14 CB676595	CB676595 OSJNEB12O
C 7	265.6	11.0	889	14 CB648876	CB648876 OSJNEB12H
C 8	265.4	11.0	864	14 CB6439712	CB6439712 OSJNEB13M
C 9	265.2	11.0	867	14 CB643557	CB643557 OSJNEB04F
C 10	265.2	11.0	867	14 CB651503	CB651503 OSJNEB16H
C 11	265	11.0	860	14 CB658123	CB658123 OSJNEC13O
C 12	263.8	10.9	891	14 CB651685	CB651685 OSJNEB16L
C 13	263.8	10.9	893	14 CB648470	CB648470 OSJNEB16L
C 14	263.6	10.9	868	14 CB673619	CB673619 OSJNEB08D
C 15	263	10.9	858	14 CB647704	CB647704 OSJNEB10L
C 16	262.6	10.9	857	14 CB677847	CB677847 OSJNEB15D
C 17	262.6	10.9	868	14 CB673025	CB673025 OSJNEB07F
C 18	262.2	10.8	855	14 CB649386	CB649386 OSJNEB13E
C 19	262.2	10.8	907	14 CB643494	CB643494 OSJNEB04D
C 20	262	10.8	862	14 CB649699	CB649699 OSJNEB13M
C 21	262	10.8	909	14 CB629973	CB629973 OSJNEB06K
C 22	261.8	10.8	864	14 CB650939	CB650939 OSJNEB15K
C 23	261.6	10.8	855	14 CB675774	CB675774 OSJNEB11L
C 24	261.4	10.8	851	14 CB666100	CB666100 OSJNEB12N
C 25	260.4	10.8	856	14 CB654478	CB654478 OSJNEB07A
C 26	258.2	10.7	853	14 CB644159	CB644159 OSJNEB05D
C 27	257.4	10.6	837	29 CG209714	CG209714 OGMAB587H
C 28	255.8	10.6	846	14 CB644856	CB644856 OSJNEB06E
C 29	255.4	10.6	861	14 CB650414	CB650414 OSJNEB14N
C 30	255.4	10.6	864	14 CB669223	CB669223 OSJNEB01H
C 31	255	10.5	846	14 CB641706	CB641706 OSJNEB01H
C 32	254.8	10.5	840	14 CB651026	CB651026 OSJNEB15N
C 33	254.8	10.5	845	14 CB650590	CB650590 OSJNEB15H
C 34	254.8	10.5	845	14 CB651106	CB651106 OSJNEB15P
C 35	254.8	10.5	845	14 CB651730	CB651730 OSJNEB16N
C 36	254.8	10.5	845	14 CB665372	CB665372 OSJNEB11L
C 37	254.8	10.5	847	14 CB644050	CB644050 OSJNEB05B
C 38	254.8	10.5	928	14 CB672850	CB672850 OSJNEB07B
C 39	254.6	10.5	841	14 CB676570	CB676570 OSJNEB12O
C 40	254.6	10.5	862	14 CB643227	CB643227 OSJNEB08L
C 41	253.8	10.5	845	14 CB644800	CB644800 OSJNEB06C
C 42	253.8	10.5	852	14 CB666405	CB666405 OSJNEB13E
C 43	253.2	10.5	867	14 CB651444	CB651444 OSJNEB16G
C 44	253.2	10.5	845	14 CB675728	CB675728 OSJNEB11K
C 45	253.2	10.5	865	14 CB642929	CB642929 OSJNEB03F

ALIGNMENTS

RESULT 1
AY103647
LOCUS Zea mays PCO142084 mRNA sequence.
DEFINITION AY103647
ACCESSION AY103647.1 GI:21206725
VERSION
KEYWORDS
SOURCE
ORGANISM

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

Ze mays

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZMDB:
www.zmdb.iastate.edu.

FEATURES

source

1..2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="maizedb:638378"
/db_xref="taxon:4577"
/clone_id="Maize Mapping Project/DuPont Cornsensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

ORIGIN

Query Match 16.5%; Score 400.2; DB 11; Length 2598;
Best Local Similarity 53.4%; Pred. No. 6,3e-45;
Matches 1095; Conservative 0; Mismatches 903; Indels 53; Gaps 10;

```

QY 184 CAGCTCGAGTCGTGACGAGCTCTCAGCCAGCCCGACGAGTCGTGAGCTCAGC 243
Db 185 CACTGACGAGGTGAAGCCCATGTGGGCGAGGCCCGGACCTGTGTAAGTCAG 244
QY 244 GGGTACAGCTCAACCTCCG-----TGACGTTGTGGGCGCCGCCGAAGGGGCGCAGG 297
Db 245 GGTTCACCTCTCGCCTCGCCAGGTGGCGCCGCTGCTCCGCCAAGACGCGTCGCG 304
QY 298 GTCCGCGTCCAGAACGACGACGATCGGACGCGTCGACAAAGAGGTGACCTCTC 357
Db 305 GTGCGCGTCAAGTCACAGAGAGGCGCCCGCCCGCTCAGAGTCAGAGGATC 364
QY 358 AAGGCCCAAGCTTCAAACTCG-----TTTACGAGTTCACCAAGGTTTCGTTGCTCG 411
Db 365 CTGACTGATCGCCACGCGCGACATCTACGCGCTCACACGCGGCTTGGGCGCAC 424
QY 412 GCCGACGACGACTGAGATGACGTACGCTCCAGAAAGCGCTTATGAGCACAGTC 471
Db 425 TCCACCGCGCCACCAAGACGAGCGCGCGCTTCAGTGCAGTGTCAAGCATCTCAAC 484
QY 472 TGGGCGGTAGCCGACGTCGCTCTTCAAGCTTCGACCGCGGCTCGAACAACG 531
Db 485 GCCGGA-----ATCTTCGACCGGAGGAGGACGCGCACAG 520
QY 532 CTTCGCGTGAAGTGTCTCCGCGCGCATGTGATCCGCTCAACTGCTTACCGGTGC 591
Db 521 CTGCGCTGGAAGTACCGCGCGGAGTGTGTGCGCATCAACCTCTCTCAAGGCG 580
QY 592 CACTGCGCGTCCGCTCGTCTTGTAGGCGCTCAACCACTTGTGAACCAACCGATC 651
Db 581 TACTCGGAGATCCGCTTGTGAATCTTCAAGGCGCATCAAGAGTGTCTCAACCGGTG 640
QY 652 AGCCCATGTCTCCCTTCGCGGCTTCATCTCGAGTCGAGCAACTCAACCGCTTCG 711
Db 641 AACCCCTGCTGCGCTCGGCGGACATCAACCGCTGCGGCGACTGTGCCGCTTCG 700
QY 712 TACATGCGCGGCGCATACCGGTACCCGACGTCGAAGTTACGTTTGTGACGAGGA 771
Db 701 TACATGCGCGGCTTATACAGGCGCGCCCAACGCGAGCGCTTACCGTCAAGGAGG 760
QY 772 ACCGAAAGATCATGTTTGGCGGAGGACCATCTGCTTGTGTGAGGAGAGTCGTC 831
Db 761 AAGGTGACGCGCGCGAGGCGTTCAAG--ATCGCGGCAATCAAGGCGGCTTCTTCAAG 817
QY 832 CTCGCGCGGAGAGAGGTCTGAGTCTGTGTAACGGAACGCGCTCTCCGCTTGAAGGG 891
Db 818 CTCGAACCCCAAGAGAGGCTCGCCATGTCACAGGCAAGTCGAGGAGCTCCGCGCTCGG 877
QY 892 ACCCTCACTGACAGCTCGACATGCTCTGCTCTTCTGAGGAGCTTGAAGGCTCTC 951

```

```

Db 878 GCCACCGTATGTACAGCCCAACGCTCGCCGCTGTGAGAGTCTCTCCGCGCTC 937
QY 952 ACCGTGAGGCCATGTCTCGCCACAGAGGCTGTTCGCGCGCTTATCAAGAGTCTGC 1011
Db 938 TTTGTGAGGTATGAAACGCAAGCCCAAGTACAGGACCACTGACCCCAACTGAAG 997
QY 1012 CGCCCGCACCCCGGCGAGGTGCGGCGCAACATCCGACGCTCTTTCGCGCG 1071
Db 998 CAC---CACCCGGGTCCATCGAGCCCGCGGCATATGAGGACATCTTGAATGGCAGC 1054
QY 1072 TCGTTTCCGTTGAGCAGGAGAGGTCAAGTCAAGAACGACGAGGAGCTTCTTGC 1131
Db 1055 TCCTT---CATGAGAGGCGCAAGAGTTGAACGAGTGAACCGCGTGTGAACCAAG 1111
QY 1132 CAGGACCGGTACCGCGTGGACGTCGAGTCAAGTTCCTCGGCGCGCTGTGAGAGCATG 1191
Db 1112 CAGGACGATACCGGCTCGACGTCGCGAGTGGCGGCGCGCCCGCCAGATGAGTATC 1171
QY 1192 ATGACGCGCTACTGACATCTCTCGCTGAGAAACAACAGACCGACCAACCCGCTCTC 1251
Db 1172 CGCGCGCGCACCAAGTCCATCGAGGCGAGGTCAACTCCGTGAACCAACCGGTATC 1231
QY 1252 GACGTGAGAAACAGACAGCCGCGGCAACTTCCAGGCGTGGCTGTCTGATT 1311
Db 1232 GACGTACACCGGTGGAAGGCGCTGACGCGGCAACTTCCAGGCGACCCCATGCGGTG 1291
QY 1312 TCGATGAGAAAGACGAGCTGCACTGCGCTCATCGGCAAGCTCACTCAGAGTGC 1371
Db 1292 TCCATGGAACAAGCGCGCGCTCGCCATCGCAACTCGGCAAGCTCATGTTGGGCAATC 1351
QY 1372 ACCGAGTGTCAAGCTTCCATGAACCGCGCGCTGCTTGTGCTCTGCTG---CCGAG 1428
Db 1352 TCGAGCTGTCAAGAGTTTACAAACAGCGGCTTCACTTCAACCTGCGCGGACGCG 1411
QY 1428 GACCGGTGCTCAACTATACGCGCAAGGCTTGAACATTACATCGTGTGCTTACGCTTGC 1488
Db 1412 AACCCACGCTGGAGTACGCTTCAAGGACGAGATGCGCTTCTTCTTCTGCTTC 1471
QY 1488 GAGCTGCGCACCTTCCACACCGGCTCACTCTGTCAGCGCGGACAGATGGCAAC 1548
Db 1472 GAGCTCCAGTACCTGGGAAACCCCATCAACAACGATGACGCGGAGCAAC 1531
QY 1549 CAGGCGGTCACTGCTGCTCATCTCGCGCGCGGCACTGCGAGGCAACGAGC 1608
Db 1532 CAGGACGTAATCTCTCGGCGCTGTCTTCCGCAAGAGACGCGCAAGGATGACATC 1591
QY 1609 CTTCCTCTCTGCGCTGCGACCTGTACTGACGCTCCAGGCGGTGACCTTCCGCGG 1668
Db 1592 CTGAAGCTCATGTGTGCTCACTTACATGTGTGCGCTGTGACCGCTGAGCCTGGCCAC 1651
QY 1669 ATGAGGCTCGACTTCAAGAAAGAGTTGACACCGCTTCTCCGACTCTCTCAAGACGAC 1728
Db 1652 CTGAGAGAACATTAAGGCGTGTGAAGAACACCGTGAACCGAGTGGCCAAAGAGGTG 1711
QY 1729 CTCGCACTGCGCTCGACGTCAA---CGACTTGGCGCTGAGGTCAAGAGGCGCTCAAC 1785
Db 1712 CTGACATTAAGAACCTCTGGCGAGGCTCTCCAGGCGCGCTTCAAGCAAGAGAGCTGATC 1771
QY 1786 AAGGCTTCGACGACGAGACGATGACCTTGAGCGCGGCTGSCAGACGCTTCTTCG 1845
Db 1772 AGCGCATGACCGCGAGGCGGTGTTCAGGTAGCGGAGAGCGCGGCGACGCGGCGCTG 1831
QY 1846 TACCGCA--CGGACCGCTGTGAGCTCTCTGCTCGCGCTTCAAGTCAAGCTT 1904
Db 1832 CCGTGAATGAGAAAGTGGCGCGCGCGCTGTGTGACACAGCCCTTACAGACGCGGCAAGCG 1891
QY 1905 TACTGCGCTTAAGCGGTGAAGGTCTCTGCGCGGAAGGCAATCTGCTTCAAGCGCGGA 1964
Db 1892 GAGGAGAGCGCTTCGTTCTTCCAAAGATCACAGGTGTGAAGAGAGAGCTCGCGCGGTG 1951
QY 1965 GGTGCGCAACGCTTCTGAGACGCGGCTTCTGAGAGGCGCGGCGACGCAATACCTCTC 2024
Db 1952 CTGCTCCAGGAGGTGAGGCGCGCGGTGGCGGTGCGGAGGAC--CGCCCGCTGAC 2010

```

QY 2025 GCCGCGCAGCGCGCTCTGTAAGTCTGCGGAGAGAGCTGGCTGAGAGCGCGCCG 2084
 Db 2011 GAACCGGATCGCGGACACCGCGCTTCCCGCTTACCGCTTCTGCGCAGAGAGCTCGG 2070
 QY 2085 CGGCGACGCTTTGTCGCGGTGCGAGAGAGAGCATGCGGAGCAACGCTTCGCGCATCTA 2144
 Db 2071 CTGCGTGTCTTCTGACCGCGAGAGGCTCAAGTCCCCCGCGAGAGATGCAACAAGTGT 2130
 QY 2145 CGAGCGCATCAAGAGAGCGCGCGCATCAACACGCTCTGTCAGATGCTCGCGTAAAGGCC 2204
 Db 2131 CGTCGCGATCAACCGAGCGGAGAGCTGTGAGACCCATGCTGTGAGTCTCAAGAGATGGA 2190
 QY 2205 GAGCAAGCTC 2215
 Db 2191 CGGCAAGCGCG 2201
 RESULT 2
 AY106831 2299 bp mRNA linear HTC 16-OCT-2002
 LOCUS Zea mays PC0142079 mRNA sequence.
 DEFINITION AY106831
 ACCESSION AY106831 GI:21209909
 VERSION AY106831.1
 KEYWORDS HTC.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 2299)
 Hailey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whiteside, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
 Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 Unpublished (2002)
 2 (bases 1 to 2299)
 Coe, E.H.
 JOURNAL Direct Submission
 REFERENCE Submitted (25-APR-2002) Maize Mapping Project, University of
 AUTHORS Missouri, Columbia, MO 65211, USA
 JOURNAL If you are interested in getting corresponding physical clones,
 TITLE these are publicly available from ZmDB and may be found by BLAST
 COMMENT searching at MSU, maize.map.org; ZmDB, www.zmdb.iastate.edu; TIGR,
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequence is either Virginia Maltot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
 www.zmdb.iastate.edu.
 FEATURES
 source
 Location/Qualifiers
 1..2299
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"
 ORIGIN
 Query Match 16.0%; Score 386.2; DB 11; Length 2299;
 Best Local Similarity 15.1%; Pred. No. 5e-43;
 Matches 950; Conservative 0; Mismatches 753; Indels 22; Gaps 9;
 QY 503 TCAGCGTTCGAGCGCGCTTCGAGAAACGCTTCGCTGAGGTCTCCGCGCGCATGG 562
 Db 246 TCTTCGCGAAGCGCTTCGAGCGCGCACGCTCCGCTTCGAGATTCGCGTCCGCGCATGC 305
 QY 563 TATTCGCGTCAACTGCTCAAGCGGTGCGCACTCGCGCTCGCTCTGAGG 622

Db 306 TGTTCGATCAACACCTTCTCTAGAGGCTACTTCGCGCATCCGCTTGAGATCTTGAGG 365
 QY 623 CGCTACCACTTCTTGAACACCGCATCAAGCCCATGCTCCCTTCGCGCTCAATCT 662
 Db 366 CCAATCAACCACTCATCAACACCGCGCTCA-GCGTCTCTGCGCTGCGGGGACCATCA 424
 QY 663 CGGCGTGGGAGACTTCAAGCCGCTCTGATACATCGCGCGGCGCATCAAGGTCACCCG 742
 Db 425 CGGCGTGGGAGACTTCTGCTCCGCTGTCTACATCGCGGCTCATCAAGGCTGCCA 484
 QY 743 ACGTCAAGTTCACGTTTTCGACGAGGAAACGAGATCATGTTTCGCGGAGGCCA 802
 Db 485 ACGCGAGCGCGCTACGCTTTCAGCGAGAGAGTGAACGCGCGAGAGCTTCAAGCGCG 544
 QY 803 TCTGCTTTTGTCTGAGGAGATGCTCTGAGCCCGGAGAGAGGCTGCTGTGTCA 862
 Db 545 CGGCGATGAGAGG---CGGCTTTCAGATCAACCCAGAGAGGCGCTGCGCATGTCA 601
 QY 863 ACGGAGAGCGGCTCCGCTCGATGCGAGACCTCACTGACAGATCAGATGCTCT 922
 Db 602 ACGGAGAGCTCCGTGGGCTCGGCGCTCGCGCATGCTGTCTGACGCGCAAGCTCTCG 661
 QY 923 CGCTCTCTGAGAGCGCTTTCAGCGCTCTTCAGGTGAGAGCCATGCTGCGCAGAGGCT 982
 Db 662 CGGTCTCTGAGAGCTCTGCTCCGCTCTTCTGCGAGGTGATGAGCGAAGCGCGAGT 721
 QY 983 CGTTGCGCGCTTCATCAGACAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1042
 Db 722 ACACGAGACCTTACCTTACAGCTCAAGCAC--CACCGGCGCTCATGAGAGCGCGG 778
 QY 1043 GCAACATCGGACGCTCTTTCGCGCTGCTGCTGTTGCGGTGAGACGAGAGAGGTCA 1102
 Db 779 CCATATGAGAGACATCTGTATGAGGAGCTCTT---CATGAGAGACGCGAGAGAGTGA 835
 QY 1103 AGGTAAAGAGAGAGAGGAGATTTCTTCCGAGAGACCGCTACCGGCTCGAGCGCTC 1162
 Db 836 ACGCATGAGACCGCTGCTGTAAGCGGAGAGAGAGAGTACGCGCTCGACGTCGCGC 895
 QY 1163 AGTTCCTGCGCGCGCTCTGAGAGAGATGAGAGCGCTACTCTCTCTGCTGAGA 1222
 Db 896 AGTGCCTCGGCGCGCGCGAGATGAGAGTCAACCGCGCGCGCGCGCGCGCGCGCGCG 955
 QY 1223 ACAACAG 1282
 Db 956 TCAATCTGATCAACAG 1015
 QY 1283 GGAATTCAGAGCGCGCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1342
 Db 1016 GCAACTTCAG 1075
 QY 1343 TCATCGGAG 1402
 Db 1076 ACATCGGAG 1135
 QY 1403 GCTGCTCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1459
 Db 1136 GCTGCTCTTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1195
 QY 1460 TGAATTCACATGCTCTTACAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
 Db 1196 CCGAGATGCGAG 1255
 QY 1520 CTTTCGTCAGCGCGAG 1579
 Db 1256 ACCAGCTTCAG 1315
 QY 1580 CGGCGCGAG 1639
 Db 1316 CAG 1375
 QY 1640 GCAAGCTTCAG 1699

Db 1376 CGCTGTGCGCAGCCGCTGACCTGCGGCACTCGAGGAAAGCTCAAGAGCCGCTCAAGA 1435
 Qy 1700 CGCTTCCCGGACCTCTCTCTACAGACACTCGGCACTCGGCTTCAGACTCAAGAC--- 1736
 Db 1436 GCTGCGATGAGCGGTGGCCAGAGAGTGTGACCAACGAGCTTCGGCGGCACTTCACA 1495
 Qy 1757 TTGGCTCGAGGTCAAGAAAGCGCTCAAGAGGCTTCAGAGAGAGAGAGAGAGAG--- 1813
 Db 1496 GCGGCGCTTCAG 1555
 Qy 1814 ACCCTGAGAGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1873
 Db 1556 ACTACGAG 1615
 Qy 1874 TCTGCTCTCGGCTCTGAG 1933
 Db 1616 TGAACCAAGCGCTTCGAG 1675
 Qy 1934 CGGCGAG 1992
 Db 1676 AGATCAACAGAGTTCAG 1735
 Qy 1993 TCTTGGAG 2050
 Db 1736 GCTTGGCTTCGAG 1795
 Qy 2051 TCGTGGCGAG 2110
 Db 1796 ACCGCTGTACGCTTCATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
 Qy 2111 AG 2170
 Db 1856 TCAAGTCCCGGCGAG 1915
 Qy 2171 ACCAGCTCTCTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2215
 Db 1916 TGAACCCATGCTCTGAG 1960

RESULT 3
 AY104679 1864 bp mRNA linear HTC 16-OCT-2002
 LOCUS Zea mays POOL42078 mRNA sequence.
 DEFINITION AY104679
 ACCESSION AY104679.1 GI:21207757
 VERSION
 KEYWORDS
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 1864)
 Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S., Arthur, L.W., Hainey, M., Morgante, M. and Tingey, S.V.
 Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes
 Unpublished (2002)
 2 (bases 1 to 1864)
 Coe, P.H.
 Direct Submission
 Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
 If you are interested in getting corresponding physical clones, these are publicly available from ZmB and may be found by BLAST searching at MSL, maizemap.org; ZmB, www.zmbl.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmB: www.zmbl.iastate.edu.
 Location/Qualifiers
 1..1864
 /organism="Zea mays"
 /mol_type="mRNA"

/db_xref="MaizeDB:638373"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/Dupont Consensus Library"
 /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

Query Match 12.0%; Score 289.2; DB 11; Length 1864;
 Best Local Similarity 57.8%; Pred. No. 7.8e-30;
 Matches 594; Conservative 0; Mismatches 423; Indels 11; Gaps 4;
 Qy 662 TCCCTTCGCGGCTTCATCTGCGGTGGGCGACCTGAGCCGCTTCTGTAATGCGG 721
 Db 1 TCCCGCTCGGGGCACTACCGGCTGGGCGACCTGCTCTCAATGATGCGCG 60
 Qy 722 GCGCATACCGGTACCCCGACGTCAGAGTTCAAGTTTGGACAGAGAGAGAGAGA 781
 Db 61 GCTCATACAGGCGCCGCAAGATCC--GTGCGGTGGGCGCCGAGAGAGAGTGG 117
 Qy 782 TCATGTTTGGCGGAGGCACTCTGCTTTTGGTTCAGAGAGAGTCTGAGCCGA 841
 Db 118 GCGCGCGAGGCGTTTGAATCCCGCATCGAGACGGATTTCTTGAAGCTGAGCCCA 177
 Qy 842 AGAGAGGTCTGGTCTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
 Db 178 AGAGAGGCTCGCATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 237
 Qy 902 TCGACAGTCAAGATGCTCTGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961
 Db 238 TTTTGAAGAGCAAGTTCTGCGCTGATGAGAGAGATCTCGGGGTGTTTGGAGAG 297
 Qy 962 CCATGATGCGCAG 1021
 Db 298 TCATAGCGGCAAGCCGAGTTCACGAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAG 354
 Qy 1022 CGGCGAGTCAAGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
 Db 355 CGGAGCATAG 414
 Qy 1082 TTGAGCAGAGAGAGAGAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
 Db 415 AGCTGGCCMAAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472
 Qy 1142 ACCGCTCGAGAGTGGCTGAGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
 Db 473 AGCGCTCGAGAGTCTCTGAG 532
 Qy 1202 ACTGAGTCTCTGCTGAG 1261
 Db 533 CCAAGTCATAGAGAGAGAGAGTCACTCCGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592
 Qy 1262 ACAAG 1321
 Db 593 GTGGCAG 652
 Qy 1322 AGACAG 1381
 Db 653 ACACCGGCTGAG 712
 Qy 1382 TCAAGGCTGAG 1438
 Db 713 TGAAGCATATACAAAG 772
 Qy 1439 TCAACTATACAG 1498
 Db 773 TGAAGTGGGTTTCAAG 832
 Qy 1499 ACCTTGGCAAGCGGCTCACTACTTGTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1558

Db 833 TCCTGGCAACCCGGTGAACAACACGTCACAGAGCGGAGACAGACAACAGAGACGTGA 892

QY 1559 ACTCCGTCCCTCTCATCTCCCGCGCCGCCCATCTGCCAGAGCCCAAGACGTCTCTCTCC 1618

Db 893 ACTCCGTGGGCTCATCTCTGTCAGAAAGACCGCGAGGACGTGACATCTTGAAGCTGA 952

QY 1619 TTCTGCTTCGACCTGTACTGACAGCTCCAGCGCCCTGACCTCCGCGCATGAGCTCG 1678

Db 953 TGACGTCCACGTTCTCTGATCGCGCTGTGTCAGAGCCATCGACCTGGGCACTTCAGAGAGA 1012

QY 1679 ACTTCAAG 1686

Db 1013 ACGTGAAG 1020

RESULT 4
CB632830

LOCUS CB632830

DEFINITION OS1EB11F05.f OS1EB Oryza sativa (indica cultivar-group) CDNA

ACCESSION CB632830

VERSION CB632830.1 GI:29627819

KEYWORDS EST.

SOURCE Oryza sativa (indica cultivar-group)

ORGANISM Oryza sativa (indica cultivar-group)

REFERENCE 1 (bases 1 to 859)
Jantaasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G. Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

JOURNAL PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Place: 11 row: P column: 05
Seq primer: gta aaa cga cgg cca gtc.
Location/Qualifiers
1. 859
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OS1EB11F05"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_1b="OS1EB"
/notes="vector: pBluescript II KS +; Site_1: EcoRI; Site_2: XhoI; 24 hrs after inoculation with Rice Blast (P06-6-3)"

ORIGIN

Query Match 11.1%; Score 268; DB 14; Length 859;
Best Local Similarity 59.2%; Pred. No. 66-27;
Matches 514; Conservative 0; Mismatches 345; Indels 9; Gaps 3;

QY 550 CGCGGCGCATGTCATCGCGCTCACTGCTCAGCGCTGCCACTGCGCGCTCGCGCTC 609

Db 1 CGGCGCGCATGTCGTCGCGCATCAACCCCTCTCCAGGCGCTACTCGCGCATCGGCTC 60

QY 610 GTGTCCTTGAGCGCGTCAACCACTTCTTGAAACCAACCGCATCAAGCCCATGTCCCTC 669

Db 61 GAGATCTCTGAGGCCATCAACCAAGCTGTCTCAACACCGGCGTCAAGCCGCTGCTCCGCTC 120

QY 670 CGCGGCTCATCTCGGCGCTCGGCGGACCTCAAGCCGCTCTGTACATCGCGGCGCATC 729

Db 121 CGTGACACATCAACCGCTCCGCTGACCTGTCTCCCTGTCTACATTTGCCGCGCTATC 180

QY 730 ACCGCTCAACCCGACGTCAAGATTCAAGTTTGACAGAGGAACCGAAGATCATGTTT 789

Db 181 ACCGCGCCGCCCAACGCGCAGG---CCATCTGCGCGACAGCGAGGAAGTGAAGCGCC 237

QY 790 GCGCGGAGGCGCATCTGCTCTTTGTGTCTGAGGACAGTCGCTCGGCGCGAAGAGGAGT 849

Db 238 GAGGCGTTCAAGCTGCGCGGATCAAGAGTGTCTTTCAAGCTGAACCCCAAGAGAGT 297

QY 850 CTGCTGTGTCAACGAAACGAGCGCTCTCCGCTCGATGAGCGACCTCAATCTGACGAC 909

Db 298 CTGCGCATCTGTCATAGCAGCTCGTGGGAGCGGCGCTCGGCGCACCGATGTTGAC 357

QY 910 TTGCACATGCTCTGCTGCTCTCGACAGGCTTGAAGGCTCTCAAGGTGAGAGGCGCATG 969

Db 358 GCAACATCTGCGCGCTCTCTGCGAGGTGCTCTGGGCGGTCTTCGAGGTGATGAAC 417

QY 970 GGCACAGAGGCTCGTTCGCGCCGCTTCATCCACGACGCTTCCCGCACCCCGCGCAG 1029

Db 418 GGCAGAGCGGAGTACACGACCACTGACCAAGCTGAAGAC---CACCTGGGTG 474

QY 1030 GTGAGGTGCGCGGACATCCGACGCTCTTTCCGCTGTGTTGGCGTTGAGCAC 1089

Db 475 ATCGAGGCGCGCGCATATGAGACATCTTCGCGGAGCTGTT---CATAGCCAC 531

QY 1090 GAGAGAGAGTCAAGTCAAGAGACGACGAGGAGGCAATTCTTCCGACGAGCGTACCCGCTC 1149

Db 532 GCCAAGAGTGAAGAGATGACCCGCTGCTGAGCCGAAGAGACAGATGACGCTC 591

QY 1150 GCGAGTGCCTCAAGTCTCTGCGCGCTCTGTGAGGACATGATGACAGGCTTCACT 1209

Db 592 CCAGCTGCGCGCACTGCTGCGCGCGAGATCGAGTCACTCCGCGCACCAAGTCC 651

QY 1210 CTCTGCTCGAGAAACAACGACGACGACGACGCGCTCTCCGACGTGAGAACAGACAG 1269

Db 652 ATCGAGCGGAGTAACTCTGTAACAACCCGGTATGACGCTCACCGCGGACAG 711

QY 1270 ACCGCGCAACGCGGCACTTCAGGCGCTGTCTGATTTGATGAGAGACGAG 1329

Db 712 GCGCTCAACGCGGCACTTCAGGCGCACCCCATCGGTGTCTCATGAGCAAGCGCGT 771

QY 1330 CTGCGACTGCGCTTATGCGGAACTCACTTCAACGACGACGAGTGTCTCAAGCT 1389

Db 772 CTGCGCATGCGCAACATGGAAGCTCATGTTCCGCGAGTCTCGAGCTCTGAAAGAA 831

QY 1390 GCCATGAACCGCGGCTGCTGTCG 1417

Db 832 AGTCTACACACGCGCTGACCTCCAAC 859

RESULT 5
CB645202

LOCUS CB645202

DEFINITION OSJNB06M01.r OSJNB Oryza sativa (japonica cultivar-group) CDNA

ACCESSION CB645202

VERSION CB645202.1 GI:29640193

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

REFERENCE 1 (bases 1 to 866)
Jantaasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G. Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea

JOURNAL
COMMENT

Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: <http://genome.arizona.edu>
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 06 row: M column: 01
Seq primer: gga aac agc tat gac cat g.

FEATURES

Location/Qualifiers

1. 866
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSUNB06M01"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSUNB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"

ORIGIN

Query Match 11.0%; Score 266.8; DB 14; Length 866;
Best Local Similarity 59.7%; Pred. No. 8.7e-27;
Matches 503; Conservative 0; Mismatches 332; Indels 9; Gaps 3;

832 CTCGCGCCGAGAGAGGGTCTCGGTCTGATCAACGAGACGGCGTCTCGCTCGATGCGG 891
865 CTGAACCCCAAGAGAGTCTCGCCATCGCATGAGCAGTCCGTGGGTGCGCGCTCGCG 806
892 ACCCTAGTCTGACAGACTCGACATGCTGCTGCTGCTGCTGCAAGCGCTTGAACGGCTC 951
805 GCCACCTGATGTTGACACGCCAACATCCTCGCGCTGCTGCTGCTGCAAGTGGCTCTCGCGG 746
952 ACCTGAGAGCCATGCTGCGCCAGAGAGGCTGTTGCGCGCTTATCCAGACGCTGTC 1011
745 TTCTGAGAGGTATGAACGAGAGCGGAGTACACGACACACCTGACCAACAGCTGAG 686
1012 CGCCCGACCCCGCGCGAGTCAAGTTCGCGGCAATCCGACGCTCTTTCCGCGTCG 1071
685 CAC--CACCCTGGGTGATGACAGCGCGCGCCATCATGAGACATCTCGCGGAGC 629
1072 TCCTTGCCTTGAAGACAGAGAGAGGTCAGGTCAAGTCAAGACGACGAGGCTTCTGCG 1131
628 TCCTT---CATGAGCAGCCCAAGAGAGTGAACGAGTGAACCGCTGCTGAACCGAAG 572
1132 CAGGACCGCTACCCCTCCGACAGTGGCTCACTGTTCTCGGCGCGCTGCTGAGAGATG 1191
571 CAGGACAGGTAGCGCTCCGCAAGTCCGCGAGTGGCTCGGCGCGCAAGATCGAGTATC 512
1192 ATGACAGCCTACCTGCACTCTCTGCTCGAGAACACAGACGACGACACCGCTCTC 1251
511 CGGCGCGCACCAAGTTCATCGAGCGCAAGTCAACTCGTGAACGACACCGCTGATC 452
1252 GACGTGAGAACAGACGACCGCGCAAGCGCGCAACTTCCAGGCGTGGCTGTCTGATT 1311
451 GACGTCCACCGCGGCAAGCGCTCCACGCGCGCAACTTCCAGGCGCGCTGATG 392
1312 TCGATGAGAACAGAGTTCGACTCGCTCGCTCACTCGGCTCACTCGGCAAGTCAACTTCA 1371
391 TCCATGAGCAAGCGCGCTTCCGCTGCTGCAACATCGGCAAGTCTGATGTCGCAAGTTC 332
1372 ACCGAGTTCCTCAACGCTGCGCATGAACCGCGCGCTGCTGCTGCTGCTG---CCGAG 1428
331 TCCGAGCTCGTGAACGAGTTCACAAACAGGCGCTGACCTCCAACTGCGCGGAGCGCG 272

QY 1429 GACCCGTGCTCACTATCAGGCAAGGGCTTGACATTCATCGTGTACGCTTCG 1488
DB 271 AACCCGAGTTGATTAAGGGTTCAAGGCAACGAGATCGGAGCTCTCTACTGCTCT 212
QY 1489 GAGCTGGGCACTTTGCCAACCCCGGTGACTGCTTCTGCAAGCCCGGAGATGGGCAAC 1548
DB 211 GAGCTCCAGTACTCTGCGCAACCCCATCAACCAATGTCAGAGCGCGGAGCAAC 152
QY 1549 CAGCGGTCACTGCTGCTCTCATCTCCGCGCGCGCACTGCGGAGGCAAGCAAGCTC 1608
DB 151 CAGGACGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 92
QY 1609 CTTTCTCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
DB 91 CTCAGCTCATGACCTCCACCTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32
QY 1669 ATGAG 1674
DB 31 CTCGAG 26

RESULT 6
CB676595/c 860 bp mRNA linear EST 09-APR-2003
LOCUS OSUNB012020.1 OSUNB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION
CB676595
CB676595.1 GI:29680320
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE

JOURNAL
COMMENT

Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: <http://genome.arizona.edu>
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 12 row: O column: 20
Seq primer: gga aac agc tat gac cat g.

FEATURES

source

1. 860
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSUNB012020"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSUNB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI, Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

ORIGIN

Query Match 11.0%; Score 266.6; DB 14; Length 860;
Best Local Similarity 59.8%; Pred. No. 9.2e-27;
Matches 503; Conservative 0; Mismatches 329; Indels 9; Gaps 3;

QY	83	CCCGAAGAGAGGCTCTCGGTCTGCGTCAACGGAAACGCGCGTCTCCGCTCGATAGGAGACCT	896
Db	860	CCCCAAGGAAGGCTTCGCACTCGTAATGGCACTCGTGGAGTGGCGCTCGCGCAC	801
QY	897	CAGTCTGCACACTGACATGCTCTTGCTCTCTCGACAGGCTTGAAGGCTTCAAGGT	956
Db	800	CGTGAATGTTCCAGCGCAACATCCTCGCGTCTGTCCAGAGTGTCTCGGCGAGTCTTG	741
QY	957	GGAAGCCATGTCGCGCCACAGAGGCTGTTGGCGCGTTCATCCAGACATCTGCGGCC	1016
Db	740	CGAGGTGATGAACCGCAAGCCGAGTACACGACCACTGACCCCAAGCTGAAGCAC--	683
QY	1017	GGACCCCGGCGCAGGTGCGAGTTCGAGCGCAATCCGACACGCTCTTTCCGCTCGTGT	1076
Db	682	-CACCTTGGGTGATCCAGAGCGCGCGGCATCATGAGACACATCTCGCGGAGAGTCTGT	624
QY	1077	TGCGCTTGAGCACGAGAGGAGGTTCAGATGTCAGACACAGAGGCAATTTCCGACAGA	1136
Db	623	---CATGAGCCACCGCAAGAGGTGAACGAGATGAGACCGCGTGTGAAGCCGAGACAGA	567
QY	1137	CCGCTAACCCGCTCCGCAAGTGCCTCAGTTCTCTCGCGCCGCTCGTGAAGACATAGTCA	1196
Db	566	CAGATACGGGCTCCGCAAGTGCAGGCAAGTGTGCGCCGACAGATGAGATCATCCGCGC	507
QY	1197	CGCTACTTGCATCTCTCTGCTCGTGCAGAACAAACAGACGACCGACACCGCTTCTGAGCT	1256
Db	506	CGCGACCAAGTCCATCGAGCGCGAGGTCACTCCGTAACGACACCCCGTATGAGGT	447
QY	1257	CCAGAACAAAGCAGACCGGCGACCGGCGCACTTCCAGCGGTGGCGTGTGATTCGAT	1316
Db	446	CCACCGCGGCAAGGCGCTCCACGGGGGCACTTCAGAGGACCCCATGAGTGTCCAT	387
QY	1317	GGAGAAAGACAGGCTGCGACTCGCCTTCATCGGCAAGCTCAACTTCAGCAGATGACGA	1376
Db	386	GGACAAACGCCGCTGTGCGCATCGCCAAATCGGCAACTGATGTGCGCAGTTCGGA	327
QY	1377	GTTGCTCAACGTCGCAATGAACCGGCGCTGCTTCTGCTCGTGG---CCGAGACCC	1433
Db	326	GCTCGTGAACGAGTTCTTACAAACAAAGGGCTGACCTTCACATGGCGGAGCCGACACC	267
QY	1434	GTCGCTCACTATCAACGCAAGGGCTTGAACATTCACATGCTGCTTACGCTTCGAGCT	1493
Db	266	GACCTTGAAGCTAACGGGTTCAAGGGGACCGAGATGCGATGGCTCTTACTGTGTGAGCT	207
QY	1494	CGGCACTTGCACAAACCGGATCACTACTTGTGTCAACCCGACAGATGGGACCAACGCG	1553
Db	206	CCAGTACTCTCGGCAACCCCATCAACCAATGTCACAGCGGACCAACAAACAGAGA	147
QY	1554	CGTCAACTGCTGCGCTCATCTCCGAGCGCGCGCACTGCGAGGCGCAAGAGTCTTTT	1613
Db	146	CGTGAACCTGCTGGGTCTGTCTCGGCGCAAGGAAGACCTTGAGGCGGTGGAATCTCTCA	87
QY	1614	TCCTCTTTCGCTCGCACTGTACTGTGACGCTTCAGGCGCTGCGACCTTCGCGCATGGA	1673
Db	86	GCTCAGCACTCAACTACATGATGACCTGTGCGACGCGTGCAGACCTTGCCACACTCGA	27
QY	1674	G 1674	
Db	26	G 26	

REFERENCE	Enthartoideae; Oryzaeae; Oryza.
AUTHORS	Janteauyirarat, C., Lu, G., Gowda, M., Harfield, J., Zhou, B., Mazur, E., Kudara, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
TITLE	Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
JOURNAL	Unpublished (2003)
COMMENT	Contact: Rod Wing Arizona Genomics Institute University of Arizona Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA Tel: 520 626 3367 Fax: 520 621 9288 Email: http://genome.arizona.edu PCR primers FORWARD: gta aac cga cgg cca gtc BACKWARD: gga aac agc tat gac cat g Plate: 12 Row: H Column: 22 Seq primer: gga aac agc tat gac cat g.
FEATURES	Location/Qualifiers
source	1. .889 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="mRNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /clone="OSUNBD12H2" /tissue_type="Leaf" /dev_stage="3 week" /lab_host="DH10B" /clone_id="OSUNDB" /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (che 86061)"
ORIGIN	
Query Match	11.0%; Score 265.6; DB 14; Length 889;
Best Local Similarity	59.8%; Pred. No. 1.3e-26;
Matches	502; Conservative 0; Mismatches 339; Indels 9; Gaps 3;
Db	838 CCGAAGAGGGTCTTCGGCTTCGTGTCACGGAACGAGCGCGCTTCGCGCTTGATGAGCGACCTC 897
Db	859 CCGAAGAGGGTCTTCGGCTTCGTGTCACGGAACGAGCGCGCTTCGCGCTTGATGAGCGACCTC 800
Qy	858 AGTCGACGAGCTCGGACATGCTTCGCTCCCTTCGAGGCGCTTGACGCGCTCAACGATG 957
Db	799 GTGATGTCGAGCGCAACATCTTCGCGCTTCGTCGAGAGTGCTCTCGGCGGTTCCTGCG 740
Qy	958 GAGGCGATGTCGAGCCAGACAGGGGCTGTTGCGCGGCTTCATCCAGACGTCGCGCGCC 1017
Db	729 GAGGTGATGAGACGCGACAGCCGGAATACACCGACCACTGACCCCAAGCTGMAACAC-- 683
Qy	1018 CACCCGCGCAGGTGCGAGTGGCGCGGACAATCCGCAAGCGCTTCGCGCTTCGCGCTT 1077
Db	682 CACCTGGGTGATGAGGCGCGCGCGCTTCATGAGACATCTTCGCGCGGAGCTCTT- 624
Qy	1078 GCCGTTGAGCAGAGAGAGAGTCAAGGTCAAGAGCAAGAGGCGATCTTTCGCGAGAC 1137
Db	623 --CATGAGCCACGCGCAAGAGTGAACAGATGAGACCGCGTGTGAAGCCGAGACAGAGAC 566
Qy	1138 CGTACCCGCTTCGAGAGTCCGCTCACTTCTTCGCGCGCGCTCGAGAGGAGATATGAC 1197
Db	565 AGTATCGGCTTCGACAGTGGCGAGTGGCTTCGCGCGCGCGAGATGAGATATCGCGCC 506
Qy	1198 GCGTACTGCACTCTCGCTCGAGAACACAGACGACCGCAACCCGCTCCTCGACGTC 1257
Db	505 GCGACCAAGTCCATGAGACCGAGAGTCAACTCCGTAACGACACACCCGATGATCGACGTC 446
Qy	1258 GAGAACAAAGCAGACCCGACGCGGCGCACTTCCAGGCGTGGCTGTTCGATTTCATG 1317
Db	445 CACCCGCGCAAGGCGCTCCAGCGCGGCAACTTCCAGGCGACCCCGATGATGATGTCATG 386
Qy	1318 GAGAGACCAAGCGCTCGACCTGCGCCCTCATCGGACCAAGTCAACTTCACGACGACCGAG 1377

ORIGIN	Query Match
11.0%; Score 265.4; DB 14; Length 864;	

QY	835	GGCCCCGAGAGAGAGGCTCGCTGTCTGATCTCAACGGAAACGGCCGCTCTCGCTCGATGGACAC	894
Db	862	GACCCCAAGGAAGGCTCTGGCCATCTGTCATATGGACACGCTCGTGGGGTGGCGCTCGCGCC	803
QY	895	CTCACTCTGCACGACTGCGACATCTCTCGCTCTCTTCTGCAAGGCTTGAACGGCTCTGACG	954
Db	802	ACCGTATGTTGCAAGCCCAATCCTCGCCGCTCTGTTCCAGAGGCTCTCGGCGGTTCTC	743
QY	955	GTGAGAGCGCATGCTGGGCGACGAGGAGTCAAGTCAAGACGACGAGGACATTTCTTGGCAG	1014
Db	742	TGCGAGGTGATGAAACGGCAGCCGAGTACACGACCACTGACCCCAAGCTGAAGAC	683
QY	1015	CCGCACCCCGGCCAGGTGAGTGGCGCGCAATCCGCAACGCTCTTCCGGCTCTGCG	1074
Db	682	---CACCTTGGGTTGATGAGGCGCGCCGACATGAGACATCTCGCCGGGAGCTCG	626
QY	1075	TTTTCGCTTGAAGCAGAGAGAGAGTCAAGTCAAGACGACGAGGACATTTCTTGGCAG	1134
Db	625	TT---CATGACCAACGCGCAAGAGTGAACGAGTGAACCCGCTGCTGAAGCCGAACAG	569
QY	1135	GACCGCTAACCCGCTCGCCAGCTCGCTCAGTTCTCGGCCCGCTCGTGAAGGACATGATG	1194
Db	568	GACAGGTACGGCTCTCCGACGTTGGCCGACGTGGCTGGCCCGGACATCGAGGTATCCG	509
QY	1195	CACGCTTACTGACTCTCTGCTGTGAAGACAACAGACGACCGAACCCGCTTCTTGAC	1254
Db	508	GCCGCCACCAAGTCAATGACGCGGAGTCAACTCCGTGAACGACAAACCCGAGTATCGAC	449
QY	1255	GTCCAGAAACAAGCAGACCGCGACCGGCGCAACTTCCAGGCGTGGGCTGTCTGATTTGG	1314
Db	448	GTCCACCGCGGCAAGGCGCTCCACGCGGCACTTCCAGGAGCACCCCAATCGGTGTCTC	389
QY	1315	ATGAGAAAGACCAAGGCTCGACTGCGCCCTATCGGACAGCTCAACTTCAACGAGTGCAC	1374
Db	388	ATGAGACAACGCCCGCTCGCCATCGCCCAATCGGCAAGCTCAATGTTCCGCGAGTTCTCC	329
QY	1375	GAGTTGCTCAACGCTGCGATGAACCGGGCGCTGCTCGTCCGCTG---CCGAGGAC	1431
Db	328	GAGTTCGTGAACGAGTTCTTCAACACAGGCTGACCTTCAACTGCGCGGACCCGAC	269
QY	1432	CCGTGCTCAACTATCAACGCGCAAGGCTTGACATTCAATCGTCTCTTAACGCTTGGAG	1491
Db	268	CCGAGCTTGAACTAACGGGTTCAAGGGGACCGAGATGCCATGCGCTCTCACTCTCTGAG	209
QY	1492	CTGGGCAACCTTGGCAACCCGGGTCACTACTTGTGTCAAGCCGCGCAAGATGGGCAACGAG	1551
Db	208	CTTCAGTACTTCGCAACCCCATATCAACCATATGTCCAGAGCGCGGAGACACAACACAG	149
QY	1552	GCGCTCAACTGCTGCTCTCATCTCCGCGCGCCGACCTGACGAGGSCCAACGACGTCTT	1611
Db	148	GAGGTGAATCTGCTGGGTCTCGTCTCGCCAGGAAGACCTCGAAGCGGTGGAATCTCTC	89
QY	1612	TCTCTCTTCTGCGCTTCGCACTCTGTACTGAGCTTCAAGGCGTTCGACCTTCGCGGAGAG	1671
Db	88	AAGCTATGACCTTCACTCATCTGTGCGCTGTGCGAGCGGTGACCTTGGCCACCTTC	29
QY	1672	GAG 1674	
Db	28	GAG 26	

RESULT 9

CB643557/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORIGIN

ORIGIN

867 bp

OSUNB04F02.r

OSUNB

OSUNB04F02 3', mRNA

GI:29638548

Oryza sativa (japonica cultivar-group)

mRNA

linear

EST 08-APR-2003

(japonica cultivar-group)

CDNA

ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrharioideae; Oryzaceae; Oryza.

REFERENCE
1 (bases 1 to 867)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)

TITLE
Unpublished (2003)

AUTHORS
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

JOURNAL
COMMENT
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: F column: 02
Seq primer: gga aac agc tat gac cat g.

FEATURES
source
1..867
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB04F02"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (che
86061)"

ORIGIN
Query Match 11.0%; Score 265.2; DB 14; Length 867;
Best Local Similarity 59.6%; Pred. No. 1.4e-26;
Matches 504; Conservative 0; Mismatches 333; Indels 9; Gaps 3;

Db 832 CTGGGCCCCAAGAGGCTCTCGTCTGCTCAACGGAAGCGCGCTCCGCTCGATGGCG 891
Db 865 CTGAACCCCAAGAGAGGTCTGCCATCTGTCATGCGACGTCCTCGGGGTCCGCTCGCG 806
Oy 892 ACCCTCAGCTCGACGACTCGACATGCTCTGCTCTCTCGACAGGCTTGACCGCTCTC 951
Db 805 GCCACCGTATGTTCCAGCGCAACATCTCTCGCGCTCTGTCGAGGTGCTCTCGCGCGTG 746
Oy 952 ACCGTGAGGCGCATGCTCGGCGACGAGGCTGTTGCGCGCGCTTCATCGACGCTCTGC 1011
Db 745 TTCTGCGAGGTATGTAACGCGCAAGCGGAGTACACCGAACCACTGACCAACACTGAG 686
Oy 1012 CGCGCGCCACCGCGGACAGGTGAGGTGCGCGGCAATCCGACGCTCTTCGCGCTCG 1071
Db 685 CAC--CACCTCGGTGCTCATTCAGGCCCGCGCATATGAGACATCTCTCCGCGAGC 629
Oy 1072 TCGTTTCCGTTGAGCAGAGAGAGAGTCAAGTCAAGAGAGAGAGGATTTCTTCCG 1131
Db 628 TCGTT---CATGAGCCAGCCCAAGAGTGAAGAGATGAGATGAGACCCGCTGTAACCGGAG 572
Oy 1132 CAGGACCGCTACCGGCTCGGACGTCGCTCACTCTCTCGCGCGCTCGTGAAGAGCATG 1191
Db 571 CAGGAGAGGTACCGGCTCGGACGTCGCGGAGTGTGCTGCGCGCATGAGAGTCTTC 512
Oy 1192 ATGACGCGCTACTCGACTCTCTGCTCGAGAAACAACAGACGACCGACACCGCTCTC 1251
Db 511 CGGCGCGCACCAAGTCAATCGAGCGCGGAGTCAATCTCGTGAAGACACACCGGTGATC 452
Oy 1252 GAGCTGAGAGAACAGAGACCGGCGACGCGCGCAATTCGAGGCTCGGCTGTCTCAAT 1311

Db 451 GAGCTGCACCGCGGCAAGAGCGCTCCACGCGGCAACTTCAGAGGACACCCCATCGGTG 392
Oy 1312 TCGATGAGAAAGACAGAGCTCGACCTCGACCTCATCGGCAAGCTCACTTACGCAAGTGC 1371
Db 391 TCGATGAGAAAGCGCCGCTCTGCGCATCGCCAAACATCGGCAAGCTCATGTTGCGGCAATT 332
Oy 1372 ACCGATGCTCAACAGCTGCCATGAACCGGCGCTGCTTGTCCTGCTG---CCGAG 1428
Db 331 TCCGAGCTGTAACAGATTCTACACACAGGCGGTGACCTTCACACCTCGGCGGAGCGCG 272
Oy 1429 GACCGCTGCTCAACTATACAGGAGCGGCTTGAGCAATTCAATCGCTGTGCTTGGCTTCG 1488
Db 271 AACCGAGCTTGAAGTACATGAGGTTCAAGGACACGAGATCGCATGCTCTTACTGCTCT 212
Oy 1489 GAGCTGCGCAACTTGTCCCAACCGCGGTCACTACTCTGTCGACCGCGGAGAGATGGCAAC 1548
Db 211 GAGCTCGAGTACCTGCGCAACCCCATCAACCATATGTCAGAGCGGAGAGAGCAAC 152
Oy 1549 CAGGCGCTCACTGCTGCTCTCATCTCCGCGCGCGCGCATCGCGGAGGCGCAAGAGCTC 1608
Db 151 CAGGACGTGAACCTGCTGGGTCTGCTCTGCGCGCAAGAACCTTGAGGCGGTGAGATC 92
Oy 1609 CTTTCTCTCTTTCGCGCTCGACCTGACTGACAGGCTTCGAGCGCGCTCGACTCCGCGCG 1668
Db 91 CTCAGCTCATGAGCTCCACTCATCATGTCGCTGCGCGAGCGCTCGACCGTCCGAC 32
Oy 1669 ATGAG 1674
Db 31 CTCGAG 26

RESULT 10
CB651503/c 867 bp mRNA linear EST 08-APR-2003
LOCUS OSJNB16H19.1 OSJNB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION OSJNB16H19.1 3', mRNA sequence.
ACCESSION CB651503
VERSION CB651503.1 GI:29646496
KEYWORDS EST
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrharioideae; Oryzaceae; Oryza.
1 (bases 1 to 867)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)

TITLE
Unpublished (2003)

AUTHORS
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu

JOURNAL
COMMENT
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 16 row: H column: 19
Seq primer: gga aac agc tat gac cat g.

FEATURES
source
1..867
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB16H19"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"

ORIGIN

/clone_11b="OSJNEB"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 24 hrs after inoculation with Rice Blast (Che
 86061)"

Query Match 11.0%; Score 265.2; DB 14; Length 867;
 Best Local Similarity 59.7%; Pred. No. 1.4e-26;
 Matches 504; Conservative 0; Mismatches 333; Indels 9; Gaps 3;

832 CTGCGCCCGAAGAGGGTCTCGCTGCTCAACGAGACCGCGTCCGCGCTCGATGAGC 891
 865 CTGAACCCCAAGAGAGTCTCGCATCTGATAGACACGCTCGGGGTGGCGCTCGCG 806
 892 ACCCTGAGTGCAGACTGCAATGCTCTGCTCTCTCGAGGCGCTTGAAGCGCTCTC 951
 805 GCCACCGTGAATGTTGACGCGCAACATCTCGCGCTCTGTCGAGAGTCTCTCGCG 746
 952 ACAGTGAAGGCGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011
 745 TTCTGCGAGGTGAATGAACGCAAGCGAGTACACCGACCTTACCCCAAGCTGAAG 686
 1012 CGCCCGGACCCCG 1071
 685 CAC--CACCTGGGTGATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 629
 1072 TCGTTTCCGTTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1131
 628 TCGTT--CATGAGCCACGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 572
 1132 CAGGACCGCTACCGCGCTCGCGCAAGTGGCTTCTGCGCGCGCGCGCGCGCGCGCG 1191
 571 CAGGACGAGTAAAGCGGCTCGCGCAAGTGGCTTCTGCGCGCGCGCGCGCGCGCG 512
 1192 ATGACGCGCTTACGACTCTCTGCTCGGAAACAACAGACGACCGCAACCGCTCTC 1251
 511 CG 452
 1252 GACGTGAGAAACAAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1311
 451 GAGTTCACCG 392
 1312 TCGATGAGAAACAAGCGCTCGCGCAAGTGGCTTCTGCGCGCGCGCGCGCGCG 1371
 391 TCCATGAGCAACCGCGCTCGCGCAAGTGGCTTCTGCGCGCGCGCGCGCGCG 332
 1372 ACCGAGTGGCTCAAGCTGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1428
 331 TCCGAGCTGCTGAACAGTCTTCAACAACGCGGCTGACCTCCACCTGCGCGG 272
 1429 GACCGCTGCTCAATATACGAGCGGCTTGAACATTCATGCTGCTTACGCTTGC 1488
 271 AACCGAGCTGAGACTAGGGGTTCAAGGCGACCGAGATGCGCATGCTCTTATG 212
 1489 GAGCTGAGCGCACTTCCCAACCGGCTACTACTTCTGTCAGCGCGCGAGAGTGG 1548
 211 GAGCTCAAGTACTTCCCAACCGGCTACTACTTCTGTCAGAGGCGGAGACACA 152
 1549 CAGGCGCTCAACTGCTGCTGCTCAATCTCGCGCGCGCGCGCGCGCGCGCGCG 1608
 151 CAGGAGTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 92
 1609 CTTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1668
 91 CTCAGACTCATGACCTCGACCTAATCTGCTGCTGCTGCTGCTGCTGCTGCT 32
 1669 ATGAG 1674
 31 CTCGAG 26

RESULT 11
 CB658123/c

LOCUS CB658123 860 bp mRNA linear EST 09-Apr-2003
 DEFINITION OSJNEC13022.r OSJNEC Oryza sativa (japonica cultivar-group) cDNA
 ACCESSION CB658123
 VERSION CB658123.1 GI:29661848
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 860)
 Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
 Kudrna,D., Dean,R., Soderlund,C., Wang,R. and Wang,G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished (2003)
 CONTACT: Rod Wing
 JOURNAL Arizona Genomics Institute
 COMMENT University of Arizona
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
 85721-0088, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR PRIMERS
 FORWARD: gta aac cga cga cga gtc
 BACKWARD: gga aac agc tat gac cat g
 Plate: 13 row: 0 column: 22
 Seq primer: gga aac agc tat gac cat g.
 FEATURES
 source 1..860
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSJNEC13022"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_11b="OSJNEC"
 /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
 XhoI; 6 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN

Query Match 11.0%; Score 265; DB 14; Length 860;
 Best Local Similarity 59.7%; Pred. No. 1.5e-26;
 Matches 502; Conservative 0; Mismatches 330; Indels 9; Gaps 3;

837 CCCGAGAGGAGTCTCGTCTGCTCAACGAGACCGCGTCCGCGCTCGATGCGACCT 896
 860 CCCAAGAGAGTCTCGCATCTGATAGGAGCTCGTGGGTCGCGCGCGCGCGCG 801
 897 CAGTTCAGAGCTCGCATGCTCTGCTCTCTGCGAGGCGCTTGAAGGCTTCAAG 956
 800 CCGTATGTTGAGCGCAACATCTCGCGCTCTGCTCGAGAGTCTCTCGCGGTTCT 741
 957 GAGGCGCATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1016
 740 CAGAGTATGAACCGGCAAGCGGAGTACACCGACCACTTACCAAGTGAAGAC-- 683
 1017 GACCCCGCGCGAGTGAAGTGGCGCAACATCGCGCAAGCTCTCTTCCGCTGCT 1076
 682 -CACCTGGGTGATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 624
 1077 TCGCGTTGAGCAGAGGAGGAGTCAAGTCAAGAGCAAGAGGAGTCTTCCGCGAG 1136
 623 ---CATGAGCCACCGCAAGAGTGAACGAGTGAACCGGCTCTGAAGCCAGAGAG 567
 1137 CCGCTACCGGCTCGCGACGCTGCGCTGCTCTGCGCGCGCGCGCGCGCGCGCG 1196
 566 CAGGATACCGGCTCGCGACGCTGCGCGAGTGGCTCGCGCGCGCGAGATGAGTATCGCG 507

QY 1197 CGCCTACTGACTCTCTGCTCGAGAAACAACGACGACCAACCCGCTCTCGACGT 1256
 Db 506 CGCACCAAGATCATGACGAGCGGAGTCACTCCGTGAACAACAACCGGTGATCGACGT 447
 QY 1257 CGAAGAACAGACGAGCGGCAACGCGGCAACTTTCAGAGCGTTCGGTGTCTGATTTTCAT 1316
 Db 446 CACCCCGGCAAGGCGCTCCACGCGCACTTCCAGGACACCCCATCGGTGTCTCAT 387
 QY 1317 GAGAAAGACAGGCTTCGACTCGCCTCATCGCAAGCTCACTTCACGACGATGACCGA 1376
 Db 386 GAGCAAGCGCCGCTCTGCGATCGCCAACTCGGCAAGCTCATGTTCCGCGAGTTCTCCGA 327
 QY 1377 GTTGTCTCAACGCTGCGCATGAACCGGCGCTGCTTCTGTGCTCGTG--CCGAGAGACC 1433
 Db 326 GCTCCGGAACGAGTTTCAACAACAACGCGCTGACCTTCAACCTGCGCGGACCGCAACCC 267
 QY 1434 GTGCTCACTATCAACGAGGCGCTTGAACATTCATTCGCTGCTTACGCTTTCGAGCT 1493
 Db 266 GACTTGAATCAAGGTTTCAAGGACCGAAGATCCGCTCTCTCTCTGAGCT 207
 QY 1494 CGGCACTTGTCCACCGCGGTCACTTCTGTCAGCGCCGCAAGATGCGCAACAGGC 1553
 Db 206 CCAATCTCTGCGCAACCCATCAACAACCATGTCAGAGCGGAGCAACAACAGGA 147
 QY 1554 CGTCACTGCTGCTCTCATCTCGCGCGCGCACTGCGCAAGCGCAAGTCTTTC 1613
 Db 146 CGTGAATCTGCTGAGTCTCTCTGCGCGAGAAAGCCCTGAGGCGGTGAGATCTTCA 87
 QY 1614 TCTCTTCTGCTGCGCACTTCACTGCAAGCTTCAGGCGGTGCACTTCGCGGATGGA 1673
 Db 86 GCTCATGACTTCACTCATCATGCTGCGCTGTGCGAGGCGGTGAGCTTCCGACCTCGA 27
 QY 1674 G 1674
 Db 26 G 26

RESULT 12
 CB651685/ 891 bp mRNA linear EST 08-APR-2003
 DEFINITION OSUNB16L23.r OSUNB Oryza sativa (japonica cultivar-group) cDNA
 OSUNB16L23.3, mRNA sequence.
 VERSION CB651685
 ACCESSION CB651685.1 GI:29646678
 KEYWORDS EST.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 BUKARYOA; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.
 Jantassuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
 Kudrna, D., Dean, R., Soderlund, C., Wang, R. and Wang, G.
 Large-scale identification of ESTs involved in the interaction
 between rice and Magnaporthe grisea
 Unpublished (2003)
 Contact: Rod Wang
 Arizona Genomics Institute
 University of Arizona
 Biological Sciences West, 448A, P.O. Box 210089, Tucson, AZ
 85721-0089, USA
 Tel: 520 626 3967
 Fax: 520 621 9288
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: gta aaa cga cgg cca gta
 BACKWARD: gga aac agc tat gac cat g
 Plate: 16 row: 1 column: 23
 Seg primer: gga aac agc tat gac cat g.
 Location/Qualifiers
 1..891
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"

ORIGIN
 Query Match 10.9%; Score 263.8; DB 14; Length 891;
 Best Local Similarity 59.3%; Pred. No. 2.2e-26;
 Matches 505; Conservative 0; Mismatches 337; Indels 9; Gaps 3;
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="OSUNB16L23"
 /tissue_type="Leaf"
 /dev_stage="3 week"
 /lab_host="DH10B"
 /clone_lib="OSUNB"
 /note="Vector: pBluescript II KS+; Site 1: EcoRI; Site 2:
 XhoI; 24 hrs after inoculation with Rice Blast (che
 86061)"

QY 827 TCGTCTCGGCGCCGAGAGAGGCTCGGCTGCTGATCAAGAGAACGCGCTCCGCTCGA 886
 Db 870 TCACGTGAACCCCAAGAAAGTCTCGGCATCTGTAAATGAGACGTCTGGGGTCGGGCG 811
 QY 887 TGGCACTCTAGTGTGACAGACTGCAATGCTCTGCTCTCTCGAGGCTTTAGCG 946
 Db 810 TCGCGCCACCGGTGATGTTGACGCGCAACATCTCGCGTCTGTCCGAGGTGCTCGG 751
 QY 947 CTCTACAGGTGAGGCACTGTCGCGCAGAGGCTGCTTGCAGCCCTTCAATCCAGACG 1006
 Db 750 CGGTCTTCCGAGGTATGAACGCAAGCCGAGTACACGACCACTGACCAACAAGC 691
 QY 1007 TCTGCGCGCCGACCCCGGCGAGGTGAGTTCGCGCGCAATCCGACGCTCTTCG 1066
 Db 690 TGAAGAC---CACCTGGGTGATGAGGCGCGCATCATGAGACATCTCTCCCG 634
 QY 1067 GCTGCTGTTTCCGTTGAGCACAGAGAGAGTCAAGTCAAGGACGACGAGGCAATTC 1126
 Db 633 GAGCTCTGTT---CATGAGCACGCGCAAGAGTGAAGATGATGATCCGCTGTGAAGC 577
 QY 1127 TTGCGCAGGACCGCTACCCGCTCCGACGTCGCTCACTTCTCGGCGCGCTCTGAGG 1186
 Db 576 CGAAGAGGAGAGATGACGCGCTCCGACGTCGCGCGAGTGCCTGGCGCCGAGATCGAGG 517
 QY 1187 ACATGATGAGCGCTTACTGACTCTCTGCTGAGAACAAACAACAACGACCAACCCCGC 1246
 Db 516 TCATCCGCGCCGCCCAAGTTCATGAGCGCGAGGTAACTCGTGAACAACAACCCG 457
 QY 1247 TCTGACGCTGAGAAACAGACGACCGGCAAGCGGCGCACTTCAGAGCGTGGCTGCT 1306
 Db 456 TGAATGACGTCACCGCGGCAAGGCGCTCCACGGGGAATTCCAGGGCAACCCCATCG 397
 QY 1307 CGATTTGATGAGAAAGACAGGCTCGCACTCGCCTCATGCGCAAGCTCACTTCAAGC 1386
 Db 396 GTGTGTCATGAGAACACCGCGCTCGCATGCGCAACATGCGGAAGCTCATGTTGCGGC 337
 QY 1367 AGTGCACGAGTTGCTCAAGCGTGCATGAACCGGGGCGCTCGCTCGCTCGCTG--- 1423
 Db 336 AGTTCCTGAGTCTGTGAACGAGTTTCAACAAGAGGCTACCTTCAACTTGGCGCA 277
 QY 1424 CCGAGACCCGTCGCTCACTATCAAGGCAAGGCTTGAACATTCATCGCTCTTACG 1483
 Db 276 GCGGACCCGAGCTTGAATACGGGTTCAAGGGGACGAGATGCGCATGCTCTACT 217
 QY 1484 CTTGGAAGTGGCCACTTGGCAACCCGGTCACTACTTGTTCAGAGCCGCAAGATG 1543
 Db 216 GCTGAGCTCAAGTACCTTGGCAACCCCATACCAATGTCAGAGCGGAGAGC 157
 QY 1544 GCAACGAGCGCTCAACTGCTGCTCATCTCGCGCGCGGCACTGCGAGGCAACG 1603
 Db 156 ACAACGAGACGTGAATCTGCTGGGTCTGCTCGGCGAGAAACCTTGAAGCGGAG 97
 QY 1604 AGGTCTTCTCTCTCTCTGCTGCTGCACTGCACTGCACTGCAAGCTTCACTTC 1663
 Db 96 ACATCTCAAGCTCAAGCTTCACTCACTCATGATGATGCTGCTGCGAGCGGACGACCTTC 37
 QY 1664 GGGCATGAG 1674

Db 36 GCCACCTCGAG 26

RESULT 13
CB648470/c 893 bp mRNA linear EST 08-APR-2003
LOCUS OSJNEB11N20.r OSJNEB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEB11N20 3', mRNA sequence.
ACCESSION CB648470
VERSION CB648470.1 GI:29643463
SOURCE EST.
ORGANISM Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 893)
Jantasuriyatat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
CONTACT: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 11 row: N column: 20
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1..893
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB11N20"
/issue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"

ORIGIN

Query Match 10.9%; Score 263.8; DB 14; Length 893;
Best Local Similarity 59.3%; Pred.No.2.2e-26;
Matches 505; Conservative 0; Mismatches 337; Indels 9; Gaps 3;

Db 827 TCCTCTCGGCGCCGAGAGAGGTCTCGTCTGTCACAGGAACGCGCTCTCCGCTTCA 886
870 TCACGCTGAACCCCAAGAGAGGTCTCGTCTGTCACAGGAACGCGCGCTCTCCGCTTCA 811
QY 887 TGGCGACCTCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 946
Db 810 TCCGCGGACCGGTGATTCGACGCGCAACATCTCTGCTGCTGCTGCTGCTGCTGCT 751
QY 947 CTCTACGAGTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1006
Db 750 CGGTGTTCTGCGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 691
QY 1007 TCTGCGCGCGGACCCCGGCGGAGGTGATGATGATGATGATGATGATGATGATGATG 1066
Db 690 TGAAGAC--CACCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 634
QY 1067 GCTCGCTGTTGCGCTTGAAGACGAGAGAGGTGATGATGATGATGATGATGATGATGAT 1126

Db 633 GGAGCTCGT---CATGAGCCACCGCAAGAGGTGAAGAGATGAGACCCGCTGCTGAGC 577
QY 1127 TTGCGCAGAGACCGCTACCCGCTTCGCAAGTGGCTTCACTTCTGCGCCGCTCGTAGAG 1186
Db 576 CGAAGCAGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 517
QY 1187 ACATGATGACCGCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1246
Db 516 TCATCGCGCGCGCCACCAAGTCATTCAGCGCGAGGTGATGATGATGATGATGATGATG 457
QY 1247 TCTTGAGCTGAGAACAGACAGACGCGCGGACGCGGACGCGGACGCGGACGCGGAC 1306
Db 456 TATGAGAGCTTCACCGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGAC 397
QY 1307 CATTTGATGAGAGAGACGAGGCTGCGACCTGCGCTTCATGCGGACGCTGACCTGACG 1366
Db 396 GTGTGTCATGAGACACGCGCTGCTGCGATGCGCAACATGCGACGATGATGATGATG 337
QY 1367 AGTGCACGAGATGCTGCTCAACGCTGCAAGAACCGCGGCTGCTGCTGCTGCTGCTG 1423
Db 336 AGTTCCTCGAGCTCGTGAACGATTTCAACAAACGCGGCTGACCTCCACCTGCGGCA 277
QY 1424 CCGAGAGACCGCTGCTCAACATGACGCGGACGCGGCTGACATTCATGATGCTGCTAG 1483
Db 276 GCGCAACCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 217
QY 1484 CTTCGAGCTGCGGACCGCTTGCACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1543
Db 216 GCTTGAGCTGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 157
QY 1544 GCAACGAGCGCTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1603
Db 155 ACAACGAGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 97
QY 1604 AGCTCTTCT 1663
Db 96 ACATCTCAAGTATATGATCTTCAACATGATGATGATGATGATGATGATGATGATGAT 37.
QY 1664 GCGCATGAGG 1674
Db 36 GCCACCTCGAG 26

RESULT 14
CB673619/c 868 bp mRNA linear EST 09-APR-2003
LOCUS OSJNE08D18.r OSJNE Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNE08D18 3', mRNA sequence.
ACCESSION CB673619
VERSION CB673619.1 GI:29677344
SOURCE EST.
ORGANISM Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 868)
Jantasuriyatat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
CONTACT: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtc

```

BACKWARD: gga aac agc tat gac cat g
Plate: 08 row: D column: 18
Seq primer: gga aac agc tat gac cat g.

```


improving the production of PAL, its analog or another optically active unnatural amino acid having PAL-like structure. The PAL polynucleotides are useful for treating a mammal having a disease, disorder or condition selected from phenylketonuria, cancer, human immunodeficiency virus infection and human cytomegalovirus infection. The present sequence represents a R. graminis PAL polypeptide

Sequence 720 AA;

Query Match 99.9%; Score 3608; DB 5; Length 720;

Best Local Similarity 100.0%; Pred. No. 1,3e-306; Mismatches 0; Indels 0; Gaps 0;

Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIIVGELLSD 60
 Db 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIIVGELLSD 60
 QY 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDKSVDFLKAQLONSVYGYTT 120
 Db 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDKSVDFLKAQLONSVYGYTT 120
 QY 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSFGLENTLPLEVVGAMVIRVN 180
 Db 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSFGLENTLPLEVVGAMVIRVN 180
 QY 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSLISAGDLSPLSYAGAITGHPDVKN 240
 Db 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSLISAGDLSPLSYAGAITGHPDVKN 240
 QY 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKREGGLVNGTAVSASMTLSLHSHMLSLLSQ 300
 Db 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKREGGLVNGTAVSASMTLSLHSHMLSLLSQ 300
 QY 301 ALTALVTEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFVEHEEVEKVKXD 360
 Db 301 ALTALVTEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFVEHEEVEKVKXD 360
 QY 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLNNNTTNDPLDVENKQTAHGNFOA 420
 Db 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLNNNTTNDPLDVENKQTAHGNFOA 420
 QY 421 SAVSISMEKTRLALIGKLNFTQCTELNMAAMNGLSPCLAAEDPSLNYHKGJLDHIA 480
 Db 421 SAVSISMEKTRLALIGKLNFTQCTELNMAAMNGLSPCLAAEDPSLNYHKGJLDHIA 480
 QY 481 AYVSEIGHLANPYTTTVOPAENGQAVNSLALISARPTAEANDVLSLLASHLYCTLOAV 540
 Db 481 AYVSEIGHLANPYTTTVOPAENGQAVNSLALISARPTAEANDVLSLLASHLYCTLOAV 540
 QY 541 DLRAMELDFKQFDPLPTLLOOHGCTGLDVNATALEVKALNKLLECTTTYYDLEPRWD 600
 Db 541 DLRAMELDFKQFDPLPTLLOOHGCTGLDVNATALEVKALNKLLECTTTYYDLEPRWD 600
 QY 601 AFSYATGIVVELLSSSPSANVTLLTAVNAKVASAEKASLSTREVRNFMQTPSSQAPAH 660
 Db 601 AFSYATGIVVELLSSSPSANVTLLTAVNAKVASAEKASLSTREVRNFMQTPSSQAPAH 660
 QY 661 YISPRTRVLXSFVRELGVQARGVFPVGOQETIGSVSRITYAIDOGKINHVLVKELA 720
 Db 661 YISPRTRVLXSFVRELGVQARGVFPVGOQETIGSVSRITYAIDOGKINHVLVKELA 720
 QY 661 YISPRTRVLXSFVRELGVQARGVFPVGOQETIGSVSRITYAIDOGKINHVLVKELA 720
 Db 661 YISPRTRVLXSFVRELGVQARGVFPVGOQETIGSVSRITYAIDOGKINHVLVKELA 720

RESULT 2

AAE27938

AAE27938 strand; protein; 720 AA.

AAE27938;

27-DEC-2002 (first entry)

Rhodotorula graminis PAL protein #1.

Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;

KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
 KW human cytomegalovirus infection; cytostatic; virucide.

OS Rhodotorula graminis.

FT Key Location/Qualifiers

FT Misc-difference 153 /label= Val, Ala

FT /note "Encoded by GYC"

PN US2002102712-A1.

PD 01-AUG-2002.

PF 24-AUG-2001; 2001US-00939408.

PR 24-JUL-2000; 2000US-00624693.

PR 24-JUL-2001; 2001WO-US023270.

PA (PCBU-) PCBU SERVICES INC.

PI Yoshida RK, Koectra AB;

DR WPI; 2002-690616/74.

DR N-PSDB; AAD45812.

PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase

PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,

PT human immunodeficiency virus or human cytomegalovirus infection.

PS Claim 26; Page 29-31; 74pp; English.

The present invention relates to Yeast (e.g. Rhodotorula) phenylalanine ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding such proteins. PAL sequences are useful for producing L-phenylalanine. They are useful for treating mammals having diseases, disorders or conditions that would benefit from treatment with PAL proteins such as phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or human cytomegalovirus (hCMV) infection. Sequences of the invention are also used in gene therapy. The present sequence is R. graminis PAL protein.

Sequence 720 AA;

Query Match 99.9%; Score 3608; DB 5; Length 720;

Best Local Similarity 100.0%; Pred. No. 1,3e-306;

Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIIVGELLSD 60
 Db 1 MAPSLDSLATTLANGFTNGSHAAPTKSAAGPTSAALRRTPGLDGHAAHQSGLIIVGELLSD 60
 QY 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDKSVDFLKAQLONSVYGYTT 120
 Db 61 PTDDVVELSGSLTVRDVVGAAKGRVRYQNDDEIRARVDKSVDFLKAQLONSVYGYTT 120
 QY 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSFGLENTLPLEVVGAMVIRVN 180
 Db 121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSXSFVSFGLENTLPLEVVGAMVIRVN 180
 QY 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSLISAGDLSPLSYAGAITGHPDVKN 240
 Db 181 SLTRGSAVRLVLEALTNFLNRITPIVPLRGSLISAGDLSPLSYAGAITGHPDVKN 240
 QY 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKREGGLVNGTAVSASMTLSLHSHMLSLLSQ 300
 Db 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKREGGLVNGTAVSASMTLSLHSHMLSLLSQ 300
 QY 301 ALTALVTEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFVEHEEVEKVKXD 360
 Db 301 ALTALVTEAMVGOQGSFAPFIHDVCRPHGOVEVANRITLLSGSSFVEHEEVEKVKXD 360
 QY 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLNNNTTNDPLDVENKQTAHGNFOA 420
 Db 361 EGLRQDRYPLRTSPQFLGFLVEDMMHAYSTLSLNNNTTNDPLDVENKQTAHGNFOA 420

```

Db      361 EGIHQDRYPLRTSPQFGLVEDMMAVSTLSIENNTTNDPLDVENKQTHAGNFOA 420
Qy      421 SAVSISMEKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKLDIHA 480
Db      421 SAVSISMEKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKLDIHA 480
Qy      481 AYASELGHLANPVTTFVOPAEQAVNSLALISARTAEANDVLSLLASHLYCTLOAV 540
Db      481 AYASELGHLANPVTTFVOPAEQAVNSLALISARTAEANDVLSLLASHLYCTLOAV 540
Qy      541 DIRAMELDFPKQFDPDLLPTLLQOHLGTLVDNALALEVKALKNKRLKEQTTTTLERPMH 600
Db      541 DIRAMELDFPKQFDPDLLPTLLQOHLGTLVDNALALEVKALKNKRLKEQTTTTLERPMH 600
Qy      601 AFSAATGTVVELSSPSANVTLLTAVNAKVAEKAISLTREVRNRFMQTPSSQAPAA 660
Db      601 AFSAATGTVVELSSPSANVTLLTAVNAKVAEKAISLTREVRNRFMQTPSSQAPAA 660
Qy      661 YLSPTTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEALDGRINHYLVKMA 720
Db      661 YLSPTTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEALDGRINHYLVKMA 720

RESULT 3
AB07694 standard; protein; 726 AA.
ID      AB07694 standard; protein; 726 AA.
AC      ABB07694;
DT      29-AUG-2003 (revised)
DT      07-AUG-2003 (revised)
DT      10-JUN-2002 (first entry)
DE      Rhodotorula pal consensu polypeptide sequence.
XX
XX      PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; cinnamate;
XX      cytosolic; anti-HIV; virucide; nootropic; dermatological; gene therapy;
XX      consensus; enzyme.
XX
XX      Rhodotorula graminis.
XX      Rhodotorula mucilaginos.
XX      Rhodotorula mucilaginos.
XX      Rhodospiridium toruloides.
XX
XX      Key Location/Qualifiers
XX      MISC-difference 1..726
XX      /note= "Xaa is any amino acid; "Xaa" indicates no
XX      consensus at that position"
XX
XX      WO200208402-A2.
XX
XX      31-JAN-2002.
XX
XX      24-JUL-2001; 2001MO-US023270.
XX
XX      24-JUL-2000; 2000US-00624693.
XX
XX      (PCBU-) PCBU SERVICES INC.
XX
XX      Yoshida RK, Kootstra AB;
XX
XX      WPI; 2002-268973/31.
XX      N-PSDB; ABA95244.
XX
XX      Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
XX      treating mammal having disease or disorder from phenylketonuria, cancer,
XX      human immunodeficiency virus infection and human cytomegalovirus
XX      infection.
XX
XX      Claim 10; Fig 2; 135pp; English.
XX
XX      The invention relates to an isolated and purified yeast phenylalanine

```

```

CC      ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC      host cell comprising the polynucleotide is useful for the production of L
CC      -PAL, by adding the polynucleotide or the host cell to a composition
CC      comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC      improving the production of PAL, its analog or another optically active
CC      unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC      are useful for treating a mammal having a disease, disorder or condition
CC      selected from phenylketonuria, cancer, human immunodeficiency virus
CC      infection and human cytomegalovirus infection. The present sequence
CC      represents a consensus polypeptide sequence of the Rhodotorula PAL.
CC      (updated on 07-AUG-2003 to correct OS field.) (updated on 29-AUG-2003 to
CC      standardise OS field)
XX
XX      Sequence 726 AA:
XX
XX      Query Match 76.7%; Score 2770; DB 5; Length 726;
XX      Best Local Similarity 77.7%; Pred. No. 3.7e-23;
XX      Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;
Qy      1 MAPSLSLATTLANGPTNGSHVAPTKSAGPTSAIRTPGLDGHAAHQSLFVQELISD 60
Db      1 MAPSLSLATTLANGPTNGSHVAPTKSAGPTSAIRTPGLDGHAAHQSLFVQELISD 60
Qy      61 P-TDVPVLSGYSLTVADVGAARGRVAVQNDDEIRAVDKSVDFLXQLONSYVGT 119
Db      61 PXTDXXELDGXSLTLDGVGAARKGRVAVXDSDEIRXIXDSVEFLRXQLONSYVGT 120
Qy      120 TGFSGSADRTEDAVSLQKALIEHOLGVTPTSXSFSVRGLENTPLELVGAMTIRV 179
Db      120 TGFSGSADRTEDAVSLQKALIEHOLGVTPTSXSFSVRGLENTPLELVGAMTIRV 180
Qy      121 TGFSGSADRTEDAVSLQKALIEHOLGVTPTSXSFSVRGLENTPLELVGAMTIRV 180
Db      121 TGFSGSADRTEDAVSLQKALIEHOLGVTPTSXSFSVRGLENTPLELVGAMTIRV 180
Qy      180 NSLTGRGSAVRLVLEALTNFNRITPIVPLRGSISASGDLPSLYAGAITGHPDVXV 239
Db      181 NSLTGRGSAVRLVLEALTNFNRITPIVPLRGSISASGDLPSLYAGAITGHPDVXV 240
Qy      240 HYLHEGTEKIMFARBAISLFLGLAVTGPXKGLVNGTAVSASAMATLSLHDSHMLLS 299
Db      241 HYLHEGTEKIMFARBAISLFLGLAVTGPXKGLVNGTAVSASAMATLSLHDSHMLLS 300
Qy      300 QALTALTYEAMVGGQSFAPPTHVCRPHQGVAVARIRITLSSGSFAVHEHEEVKVD 359
Db      301 QALTALTYEAMVGGQSFAPPTHVCRPHQGVAVARIRITLSSGSFAVHEHEEVKVD 360
Qy      360 DEGIHQDRYPLRTSPQFGLVEDMMAVSTLSIENNTTNDPLDVENKQTHAGNFOA 418
Db      361 DEGIHQDRYPLRTSPQFGLVEDMMAVSTLSIENNTTNDPLDVENKQTHAGNFOA 420
Qy      419 QASAVSISMEKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKLDIHA 478
Db      421 QASAVSISMEKTRLALALIGKLNFTQCTELNANMRGLPSCIAADPSLNHGKLDIHA 480
Qy      479 IAYASELGHLANPVTTFVOPAEQAVNSLALISARTAEANDVLSLLASHLYCTLO 538
Db      481 IAYASELGHLANPVTTFVOPAEQAVNSLALISARTAEANDVLSLLASHLYCTLO 540
Qy      539 AYDLRAMELDFPKQFDPDLLPTLLQOHLGTLVDNA---LALVYKALKNKRLKEQTTT 594
Db      541 AYDLRAMELDFPKQFDPDLLPTLLQOHLGTLVDNA---LALVYKALKNKRLKEQTTT 600
Qy      595 EPRWHAESYATGTVVELSSPSANVTLLTAVNAKVAEKAISLTREVRNRFMQTPSS 654
Db      601 EPRWHAESYATGTVVELSSPSANVTLLTAVNAKVAEKAISLTREVRNRFMQTPSS 660
Qy      655 QAPAEAYLSPTTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEALDGRINHY 714
Db      661 SSPALKYLSPTRVLYSFVRELGVQARGDVFGVQOETIGSNVSRITYEALDGRINHY 720
Qy      715 LVKMA 720
Db      721 LVKMA 726

RESULT 4

```

```

AAE27942
ID AAE27942 standard; protein; 726 AA.
XX
AC AAE27942;
XX
DT 06-AUG-2003 (revised)
DT 27-DEC-2002 (first entry)
XX
DE Phenylalanine ammonia lyase consensus protein.
XX
KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytosstatic; virucide.
XX
OS Rhodocorula graminis.
OS Rhodocorula mucilaginosa.
OS Rhodospiridium toruloides.
XX
FH Key Location/Qualifiers
FT Misc-difference /note="Yaa corresponds to unknown amino acid residue"
FT US2002102712-A1.
XX
XX 01-AUG-2002.
XX
XX 24-AUG-2001; 2001US-00939408.
XX
XX 24-JUL-2000; 2000US-00624693.
XX 24-JUL-2001; 2001WO-US023270.
XX
XX (PCBU-) PCBU SERVICES INC.
XX
XX PI Yoshida RK, Kootstra AB;
XX
XX DR WPI; 2002-690616/74.
XX
XX PT Novel isolated and purified Rhodocorula phenylalanine ammonia lyase
XX PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
XX PT human immunodeficiency virus or human cytomegalovirus infection.
XX
XX PS Claim 26; Page 47-48; 74pp; English.
XX
XX CC The present invention relates to yeast (e.g. Rhodocorula) phenylalanine
XX CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
XX CC such proteins. PAL sequences are useful for producing L-phenylalanine.
XX CC They are useful for treating mammals having diseases, disorders or
XX CC conditions that would benefit from treatment with PAL proteins such as
XX CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
XX CC human cytomegalovirus (hCMV) infection. Sequences of the invention are
XX CC also used in gene therapy. The present sequence is PAL consensus protein.
XX CC (Updated on 06-AUG-2003 to correct OS field.)
XX
XX SQ Sequence 726 AA;

Query Match 76.7%; Score 2770; DB 5; Length 726;
Best Local Similarity 77.7%; Pred. No. 3.7e-233;
Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

QY 1 MAPSLDSLATLTANGFTNGSHAAPTKSAAGPTSLRRPGLDGHAAQSOLEYVDELSD 60
DB 1 MAPSLDSITSTKXANXKXGKXPAAXASXXXXXXKXAGXKLPPTYXTQLDIVEXKXAD 60
QY 61 P-TDDVVELSGYSLTVRDVVGAAKRGKRVRYONDDETIRARVDSVDELKXQLONSVYGV 119
DB 61 PXTDXXXELDGYSLTLGDVVGAAKRGKRVRYXDSDELIRKIKDSVETLRQQLANSYRGV 120
QY 120 TGFSGSADTRTDAVSLQKALIEHQLCGVTTSSSSVSGLENTPLFVVRGAMVIRY 179
DB 121 TGFSGSADTRTDAVSLQKALIEHQLCGVTTSSSSVSGLENTPLFVVRGAMVIRY 180
QY 180 NSLTGRGSAVRLVLEALTNFNLNRTIPVPLRSISASGDLPSLYIAGATGHPDVAY 239

```

```

DB 181 NSLTGRGSAVRLVLEALTNFNLNRTIPVPLRGTISASGDLPSLYIAAITGHPDSKV 240
QY 240 HVLHEGTEKIMFAREAISLFGAEAVVLGKREGGLVNGTAVASAMATLSLDSMTLSLS 299
DB 241 HVNHEGTEKIMFAREAIALFGIEPVYLGKREGGLVNGTAVASAMATLALDHAMLSLS 300
QY 300 QALTALTVEAMYGQGSFAPFIHDVCRPHPGQVEVARNIRITLSSGFAYHEEYVKD 359
DB 301 QALTALTVEAMYGHAGSFHPFLHDVTRPHPTQIEVARNIRITLBSGXFAVHHEEYVKD 360
QY 360 DEGLIKODRYPPLRTSPQFGLVPEDMEMAYSTLSIF-NTTTPDPLDVENKOTARHGNF 418
DB 361 DEGLIKODRYPPLRTSPQFGLVPSDMIHAAVLSLEAGOSTTDNPLDIVENKXTHHGNF 420
QY 419 QASAVSISMEKTRLALAIIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHGKGDH 478
DB 421 QASAVXNTMEKTRLALAIIGKLNFTQCTELNAAANRGLPSCLAEDPSLSYHCKGDIA 480
QY 479 IAAVASELGHLANPVTTVOAPENGNAVNSLAIISARTAEANDVLSLLASHLYCTIQ 538
DB 481 AAAYTSEIGHLANPVTTVOAPENGNAVNSLAIISARTAEANDVLSLLASHLYCTIQ 540
QY 539 AVDLRAMELDFKKOPDPLPTLLOOHLGTDVNA----LALVKKALNRIEQTYYDL 594
DB 541 AVDLRAMELDFKKOPDPLPTLLOOHLGTDVNA----LALVKKALNRIEQTYYDL 600
QY 595 EPRMHDASFATGIVVELLSGSSPSANVTLTAVNAKVAABKAI SLTREVRNRFWQTPSS 654
DB 601 EPRMHDASFATGIVVELLSGSSPSANVTLTAVNAKVAABKAI SLTREVRNRFWQTPSS 660
QY 655 QAPAAVYSPRTIRVYSVREELGVQAPRGVFPVGVQCEITGSNVSRTIYAIIDGRIHY 714
DB 661 SSPALXVSPRTIRVYSVREELGVQAPRGVFPVGVQCEITGSNVSRTIYAIIDGRIHY 720
QY 715 LVKMLA 720
DB 721 LVKMLA 726

RESULT 5
AAE27943
ID AAE27943 standard; protein; 720 AA.
XX
XX AC AAE27943;
XX
XX DT 27-DEC-2002 (first entry)
XX
XX DE Yeast phenylalanine ammonia lyase protein #2.
XX
XX KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytosstatic; virucide.
XX
XX OS Rhodocorula graminis.
XX
XX FH Key Location/Qualifiers
FT Misc-difference 5 /label=Val, Leu, Phe
FT FT /note="Encoded by B7C"
FT FT Misc-difference 12 /label=Val, Leu, Phe
FT FT /note="Encoded by B7C"
FT FT Misc-difference 16 /label=Val, Leu, Phe
FT FT /note="Encoded by B7C"
FT FT Misc-difference 17 /label=Thr, Ala, Ser
FT FT /note="Encoded by DCV"
FT FT Misc-difference 19 /label=Gly
FT FT /note="Encoded by GGV"
FT FT Misc-difference 20 /label=Ser, Leu

```

FT /note= "Encoded by THG or an in-frame stop codon"
FT Misc-difference 25 /label= Thr, Pro, Ser
FT /note= "Encoded by HCC"
FT Misc-difference 27 /label= Pro, Ser
FT /note= "Encoded by YCG"
FT Misc-difference 28 /label= Ala, Pro
FT /note= "Encoded by SCW"
FT Misc-difference 34 /label= Thr
FT /note= "Encoded by AGH"
FT Misc-difference 36 /label= Arg, Ser
FT /note= "Encoded by MCM"
FT Misc-difference 39 /label= Ala, Pro, Ser
FT /note= "Encoded by BCC"
FT Misc-difference 40 /label= Arg, Gly, Trp
FT /note= "Encoded by BGG"
FT Misc-difference 48 /label= Lys, Thr, Met, Glu, Ala, Val, Gln, Pro, Leu
FT /note= "Encoded by VH3"
FT Misc-difference 54 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 56 /label= Lys, Glu, Gln
FT /note= "Encoded by VAG"
FT Misc-difference 65 /label= Glu, Asp, Val
FT /note= "Encoded by GWS"
FT Misc-difference 66 /label= Ile, Val, Leu
FT /note= "Encoded by VTC"
FT Misc-difference 76 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 87 /label= Thr, Pro, Ser
FT /note= "Encoded by HCB"
FT Misc-difference 93 /label= Asp, Ala
FT /note= "Encoded by GMC"
FT Misc-difference 102 /label= Lys, Asn
FT /note= "Encoded by AAV"
FT Misc-difference 103 /label= Arg, Ser
FT /note= "Encoded by AGB"
FT Misc-difference 109 /label= Thr, Ala, Ser
FT /note= "Encoded by DCB"
FT Misc-difference 112 /label= Asp, His, Tyr
FT /note= "Encoded by BAC"
FT Misc-difference 114 /label= Arg, Ser
FT /note= "Encoded by AGB"
FT Misc-difference 117 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 148 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 150 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 153 /label= Ile, Val, Phe
FT /note= "Encoded by DTC"

FT Misc-difference 154 /label= Glu, Asp
FT /note= "Encoded by GAB"
FT Misc-difference 157 /label= Ser, Gly, Arg
FT /note= "Encoded by VGC"
FT Misc-difference 159 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 183 /label= Thr
FT /note= "Encoded by ACB"
FT Misc-difference 223 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 225 /label= Ser
FT /note= "Encoded by TCB"
FT Misc-difference 237 /label= Ser, Thr, Ile, Gly, Ala, Val, Cys, Phe
FT /note= "Encoded by DVC"
FT Misc-difference 239 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 241 /label= Val
FT /note= "Encoded by GTY"
FT Misc-difference 242 /label= Val, Leu, Phe
FT /note= "Encoded by KTS"
FT Misc-difference 246 /label= Lys, Asn, Thr
FT /note= "Encoded by AMS"
FT Misc-difference 251 /label= Tyr, Ser, Phe
FT /note= "Encoded by THC"
FT Misc-difference 259 /label= Leu, Phe
FT /note= "Encoded by TTD"
FT Misc-difference 305 /label= Leu
FT /note= "Encoded by CTB"
FT Misc-difference 319 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 346 /label= Lys, Arg, Thr, Gln, Pro, Trp, Ser
FT /note= "Encoded by HWG or an in-frame stop codon"
FT Misc-difference 411 /label= Lys, Thr, Met
FT /note= "Encoded by AHG"
FT Misc-difference 421 /label= Thr, Ala, Ser
FT /note= "Encoded by DCC"
FT Misc-difference 457 /label= Leu
FT /note= "Encoded by CTB"
FT Misc-difference 458 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 466 /label= Pro
FT /note= "Encoded by CCB"
FT Misc-difference 487 /label= Gly
FT /note= "Encoded by GGH"
FT Misc-difference 493 /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 500 /note= "Encoded by GCH"
FT Misc-difference 518 /label= Thr

```

FT /note= "Encoded by ACB"
FT Misc-difference 536
FT /label= Val
FT /note= "Encoded by GTB"
FT Misc-difference 556
FT /label= Met, Ile, Val, Leu
FT /note= "Encoded by VTB"
FT Misc-difference 557
FT /label= Ile, Val, Leu
FT /note= "Encoded by VTC"
FT Misc-difference 558
FT /label= Thr, Ala, Pro
FT /note= "Encoded by VCB"
FT Misc-difference 559
FT /label= Thr, Ala, Ser
FT /note= "Encoded by DCG"
FT Misc-difference 561
FT /label= Ile, Leu, Phe
FT /note= "Encoded by HTC"
FT Misc-difference 562
FT /label= Lys, Glu, Gln
FT /note= "Encoded by VAG"
FT

Query Match 75.6%; Score 2730; DB 5; Length 720;
Best Local Similarity 76.5%; Pred. No. 1.2e-229;
Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

QY 1 MAPSLDGLATTLANGFTNGSHAFTSAAGPTSALETRTPGLDGHAAHOSQLEIVQELLSD 60
DB 1 MAPSDSIATISXNAGXNXXHAPKXKXTGATSLXRPXKLAPPAQTXQTDIDIXELIAD 60
QY 61 PTDDVVELSGYSLTVRDVVGAAKGRVRVQNDDEIRARVDSVDFLKAQLQNSVYGVTT 120
DB 61 PTDDXXELDGYTLTLDVVGAAKGRKRVVQTXDIDIRAKIXXVEFLKXQLXNVYVTT 120
QY 121 GFGSSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGENTLPLEVVRGAMTVRVN 180
DB 121 GFGSSADRTEDAVSLQKALIEHQLCGXLTSTXSFYLKRGLENSLPLEVVRGAMTVRVN 180
QY 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGTSASGDLSPSYIAGAITGHPDVKVH 240
DB 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGTSASGDLSPSYIAGAITGHPDVKXH 240
QY 241 VHEGTETKMPAREATSLFGLBAVLGPKEGGLVNGTAVASAMATLSHSHMLSLLSQ 300
DB 241 VHEGTETKMPAREATSLFGLBAVLGPKEGGLVNGTAVASAMATLSHSHMLSLLSQ 300
QY 301 ALTALTVEAMVQOQGSFAPFIHDVCRPHPGQVENARNTIRLLSGSSFAVEHEEVEVKXD 360
DB 301 ALTALTVEAMVGHAGSFHFLHDVTRPHPTQIEVARNTIRLLSGSSFAVHHEEVEVKXD 360
QY 361 EGI LRDRYPLRTSPQFLPVEDMMAVSTLSLENNTTTNPILDVENKQTAGNFOA 420
DB 361 EGI LRDRYPLRTSPQFLPVEDMMAVSTLSLENNTTTNPILDVENKQTAGNFOA 420
QY 421 SAVSISMETRIALALIGLNTFOCTELTANAMRGJPSCLAAEDPSLNYHGKGLDTHIA 480
DB 421 SAVSISMETRIALALIGLNTFOCTELTANAMRGJPSCLAAEDPSLNYHGKGLDTHIA 480
QY 481 AVASLIGLANVTTTFVOPAEKNOAVNSLALISAPRTAEANDVLSLLASHLYCTLQAV 540
DB 481 AVASLIGLANVTTTFVOPAEKNOAVNSLALISAPRTAEANDVLSLLASHLYCTLQAV 540
QY 541 DLRAMELDPKQDFDLPTLLQOHCTGLDYNALALEYKALNRLLEOTTITYLEPRMHD 600
DB 541 DLRAMELDPKQDFDLPTLLQOHCTGLDYNALALEYKALNRLLEOTTITYLEPRMHD 600
QY 601 AFSVATGIVVELLSSSPSANYTLTAVANAKVASAEKASISLREVRNFWOTPSQAPAH 660
DB 601 AFSVATGIVVELLSSSPSANYTLTAVANAKVASAEKASISLREVRNFWOTPSQAPAH 660
QY 661 YLSPRTVLVSFVREELGVKARGDVPFLGQDEVITIGTVNSRIYEAIKXGKXINHYVKMLA 720
DB 661 YLSPRTVLVSFVREELGVKARGDVPFLGQDEVITIGTVNSRIYEAIKXGKXINHYVKMLA 720

```

```

DB 661 YLSPRTVLVSFVREELGVKARGDVPFLGQDEVITIGTVNSRIYEAIKXGKXINHYVKMLA 720

RESULT 6
AAP83141
ID AAP83141 standard; protein; 716 AA.
XX
AC AAP83141;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 19-NOV-1990 (first entry)
XX
DE Sequence of L-phenylalanine ammonia lyase (PAL) from Rhodospiridium
DE turcoides.
XX
KM Cinnamic acid; expression vector.
OS Rhodospiridium turcoides.
XX
PN EF278706-A.
XX
PD 17-AUG-1988.
XX
PF 08-FEB-1988; 88EP-00301011.
XX
PR 06-FEB-1987; 87JP-00024705.
PR 18-JUN-1987; 87JP-00152357.
XX
PA (MITR ) MITSUI TOATSU CHEM INC.
XX
PI Fukuhara N, Yoshino S, Sone S, Nakajima Y, Makiguchi N;
XX
DR WPI; 1988-229543/33.
XX
PT Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
PT having combined promoter of tac promoter and P1 promoter for enhanced
XX expression.
XX
PS Example; Page 16-21; 37Pp; English.
XX
CC Inventors claim a recombinant plasmid contg. a DNA sequence coding for
CC PAL with the SQ in AAP80501. Procedure for cloning the structural gene
CC for PAL (AAP81101) is described as a reference example. The hybrid
CC plasmid permits more efficient expression of PAL in E.coli. PAL is used
CC for the prodn. of L-phenylalanine from cinnamic acid and ammonia (Updated
CC on 25-MAR-2003 to correct PR field.) (Updated on 27-AUG-2003 to correct
CC OS field.)
XX
SQ Sequence 716 AA;

Query Match 72.5%; Score 2619; DB 1; Length 716;
Best Local Similarity 72.1%; Pred. No. 6.2e-220;
Matches 523; Conservative 86; Mismatches 102; Indels 14; Gaps 5;

QY 1 MAPSLDGLATTLANGFTNGSHAFTSAAGPTSALETRTPGLDGHAAHOSQLEIVQELLSD 60
DB 1 MAPSLDGLATTLANGFTNGSHAFTSAAGPTSALETRTPGLDGHAAHOSQLEIVQELLSD 60
QY 61 PTDDVVELSGYSLTVRDVVGAAKGRVRVQNDDEIRARVDSVDFLKAQLQNSVYGVTT 120
DB 61 PTDDVVELSGYSLTVRDVVGAAKGRVRVQNDDEIRARVDSVDFLKAQLQNSVYGVTT 120
QY 121 GFGSSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGENTLPLEVVRGAMTVRVN 180
DB 121 GFGSSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGENTLPLEVVRGAMTVRVN 180
QY 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGTSASGDLSPSYIAGAITGHPDVKVH 240
DB 181 SLTRGSAVRLVYLEALTNFLNHRITPIVPLRGTSASGDLSPSYIAGAITGHPDVKVH 240
QY 241 VHEGTETKMPAREATSLFGLBAVLGPKEGGLVNGTAVASAMATLSHSHMLSLLSQ 300
DB 241 VHEGTETKMPAREATSLFGLBAVLGPKEGGLVNGTAVASAMATLSHSHMLSLLSQ 300

```

Db 235 VVHEGKEKILYAREAMALFNLEPVVLGPKKEGLGVNGTAVASAMATLALDAHMLSLSSQ 294
 301 ALTALTVEAMVGOQGSFAPFTIHVCRPHPGQVEYANIRITLSSGSPAVHHEEVKVKD 360
 295 SLTAMTVAMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKILLEGSRPAVHHEEVKVKD 354
 361 EGIIRKQDVPYRTSPQFGLPEVDMHMAVSTLSLE--NNTTNDNPLDVENKQTAHGNFQ 419
 365 EGIIRKQDVPYRTSPQFGLPEVDMHMAVSTLSLE--NNTTNDNPLDVENKQTAHGNFQ 414
 420 ASAVSISMEKTRALALIGKLNFTQCTELLNAANNRGLPSCLAEDPSLVNHGKGLDIHI 479
 415 AAAVANTVEKTRGLAQIGKLNFTQCTELLNAANNRGLPSCLAEDPSLVNHGKGLDIHI 474
 460 AAYASELGHLANPVTTFQPAEMGNQAVNSIALISARRTAESNDVLSLLASHLYCTLOA 539
 475 AAYTSELGHLANPVTTHQPAEMANOAVNSIALISARRTESNDVLSLLASHLYCTLOA 534
 540 VDLRAMEIDFKKQFDPPLPTLLQOHLG---TGLDV-NALALEVKKALNKLKLEQTTVDLE 595
 535 IDLRALIEFEFKKQFQPAVALVSLIDQFSGAMTGSNLRLBELVEKVKTLAKRLEQTTVDLV 594
 536 PRMHDAFSYATGTVVELLISSPSANVTITAVNAMKVASAEKALISLTREVNRRFMQTTSSQ 655
 535 PRMHDAFSYATGTVVELLISSPSANVTITAVNAMKVASAEKALISLTREVNRRFMQTTSSQ 651
 656 APAHAYLSPRTYVLSFYREELGVQARRGDVFGVQOETIGSNVSRLEYEAIKQGRINHL 715
 652 SPALSYLSRPTQILYAFVREELGVQARRGDVFLGKQEVTTISNVSKITYEAIKSGRINNVL 711
 QY 716 VKOLA 720
 Db 712 LKOLA 716
 RESULT 7
 AAB0513
 ID AAB0513 standard; protein; 716 AA.
 AC AAB0513;
 DT 25-MAR-2003 (revised)
 DT 12-NOV-1990 (first entry)
 XX
 DE L-phenylalanine ammonia-lyase.
 XX
 KM L-phenylalanine ammonia-lyase; PAL: foreign gene expression;
 culture temperature; expression regulation.
 XX
 OS Rhodosporidium toruloides.
 XX
 PN EP279665-A.
 PD 24-AUG-1988.
 PF 18-FEB-1988; 88EP-00301356.
 PR 19-FEB-1987; 87JP-00034397.
 PR 18-JUN-1987; 87JP-00152359.
 XX
 PA (MITK) MITSUI TOATSU CHEM INC.
 PI Fukushima N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y,
 DR WPI: 1988-236895/34.
 DR N-PSDB; AAN81116.
 XX
 PT Regulation of expression of foreign gene in Escherichia coli - by
 PT maintaining temp. of culture at 40 deg. C or more to suppress expression.
 PS Claim 3; Page 15-19; 36pp; English.
 CC The PAL gene is expressed in E.coli carrying a recombinant plasmid.

CC Expression is regulated by maintaining the temperature at at least 40
 CC deg. C to suppress expression. Pal is produced in high concentrations.
 CC See also EP-279664: regulation by sugar component as C source at 0.3% or
 CC more. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
 CC 2003 to correct PI field.)
 XX
 SQ Sequence 716 AA:
 Query Match 72.4%; Score 2614; DB 1; Length 716;
 Best local similarity 72.0%; Pred. No. 1,7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
 1 MAPSLDLSIATTLANGFTNGSHAPPTSAAGPTSAKRTPEGLDHAHQSLTEVQGLSD 60
 1 MAPSLDLSISHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQDIVERKMLAA 54
 61 PTDVVELSGSYSLVRVDYVGARKGRVRVONDEIRAVYDKSDVDFKQALQNSVYGVTT 120
 55 PTDVVELSGSYSLVRVDYVGARKGRVRVONDEIRAVYDKSDVDFKQALQNSVYGVTT 114
 121 GFGGSADRTEDAVSLQKALIEHQLCGVPTLSXSSFSVSGLENTLPLEVVRQAMVIRVN 180
 115 GFGGSADRTEDAVSLQKALIEHQLCGVPTLSXSSFSVSGLENTLPLEVVRQAMVIRVN 174
 161 SLTRGHSVRLVTEALTNPLNRTPIVPLRGSISASGSLSPSTAGATGHDPVYKH 240
 175 SLTRGHSVRLVTEALTNPLNRTPIVPLRGSISASGSLSPSTAGATGHDPVYKH 234
 241 VHEGTEKIMFAREASISFGLAEAVTLGPKKEGLGVNGTAVASAMATLALDAMHLSLSQ 300
 235 VVHEGKEKILYAREAMALFNLEPVVLGPKKEGLGVNGTAVASAMATLALDAMHLSLSQ 294
 301 ALTALTVEAMVGOQGSFAPFTIHVCRPHPGQVEYANIRITLSSGSPAVHHEEVKVKD 360
 295 SLTAMTVAMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKILLEGSRPAVHHEEVKVKD 354
 361 EGIIRKQDVPYRTSPQFGLPEVDMHMAVSTLSLE--NNTTNDNPLDVENKQTAHGNFQ 419
 365 EGIIRKQDVPYRTSPQFGLPEVDMHMAVSTLSLE--NNTTNDNPLDVENKQTAHGNFQ 414
 420 ASAVSISMEKTRALALIGKLNFTQCTELLNAANNRGLPSCLAEDPSLVNHGKGLDIHI 479
 415 AAAVANTVEKTRGLAQIGKLNFTQCTELLNAANNRGLPSCLAEDPSLVNHGKGLDIHI 474
 460 AAYASELGHLANPVTTFQPAEMGNQAVNSIALISARRTAESNDVLSLLASHLYCTLOA 539
 475 AAYTSELGHLANPVTTHQPAEMANOAVNSIALISARRTESNDVLSLLASHLYCTLOA 534
 540 VDLRAMEIDFKKQFDPPLPTLLQOHLG---TGLDV-NALALEVKKALNKLKLEQTTVDLE 595
 535 IDLRALIEFEFKKQFQPAVALVSLIDQFSGAMTGSNLRLBELVEKVKTLAKRLEQTTVDLV 594
 536 PRMHDAFSYATGTVVELLISSPSANVTITAVNAMKVASAEKALISLTREVNRRFMQTTSSQ 655
 535 PRMHDAFSYATGTVVELLISSPSANVTITAVNAMKVASAEKALISLTREVNRRFMQTTSSQ 651
 656 APAHAYLSPRTYVLSFYREELGVQARRGDVFGVQOETIGSNVSRLEYEAIKQGRINHL 715
 652 SPALSYLSRPTQILYAFVREELGVQARRGDVFLGKQEVTTISNVSKITYEAIKSGRINNVL 711
 QY 716 VKOLA 720
 Db 712 LKOLA 716
 RESULT 8
 AAB07693
 ID AAB07693 standard; protein; 716 AA.
 AC AAB07693;
 DT 07-AUG-2003 (revised)
 DT 10-JUN-2002 (first entry)

XX DE R. toruloides PAL polypeptide.
 XX KW PAL; yeast; phenylalanine ammonia lyase; cinnamic acid; enzyme;
 XX KM cyclostatic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
 XX OS Rhodosporidium toruloides.
 XX PN WO200208402-A2.
 XX PD 31-JAN-2002.
 XX PF 24-JUL-2001; 2001WO-US023270.
 XX PR 24-JUL-2000; 2000US-00624693.
 XX PA (PCBU-) PCBU SERVICES INC.
 XX PI Yoshida RK, Kootstra AB,
 XX DR WPI; 2002-268973/31.
 XX DR N-PSDB; ABA95243.
 XX PS Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
 XX PT treating mammal having disease or disorder from phenylketonuria, cancer,
 XX PT human immunodeficiency virus infection and human cytomegalovirus
 XX PT infection.
 XX PS Example 3; Fig 1; 135pp; English.
 XX CC The invention relates to an isolated and purified yeast phenylalanine
 XX CC ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
 XX CC host cell comprising the polynucleotide is useful for the production of L
 XX CC -PAL, by adding the polynucleotide or the host cell to a composition
 XX CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
 XX CC improving the production of PAL, its analog or another optically active
 XX CC unnatural amino acid having PAL-like structure. The PAL polynucleotides
 XX CC are useful for treating a mammal having a disease, disorder or condition
 XX CC selected from phenylketonuria, cancer, human immunodeficiency virus
 XX CC infection and human cytomegalovirus infection. The present sequence
 XX CC represents a R. toruloides PAL polypeptide. (Updated on 07-AUG-2003 to
 XX CC correct OS field.)
 XX CC Sequence 716 AA;
 XX SQ

Query Match 72.4%; Score 2614; DB 5; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY 1 MAPSLDSLTLTANGFTNGSHAPTKSAAGPTSAARRTPGLDGHAAQSLEIYQELISD 60
 DB 1 MAPSLDSLTLTANGFTNGSHAPTKSAAGPTSAARRTPGLDGHAAQSLEIYQELISD 54
 QY 61 PDDVVELSGYSLTRDVYGAARKGRVAVQNDDEIRARVDSYDFKACLOKSVYGVTT 120
 DB 55 PDDVVELSGYSLTRDVYGAARKGRVAVQNDDEIRARVDSYDFKACLOKSVYGVTT 114
 QY 121 GFGGSADRTTDAVSLQKALLHOLCGVTPTSXSSFSVSGLENTLELVVRGAMVIRVN 180
 DB 115 GFGGSADRTTDAVSLQKALLHOLCGVTPTSXSSFSVSGLENTLELVVRGAMVIRVN 174
 QY 181 SLTRGSAVRLVVLALTNFLNHRITPVPPLRGISASGDSPLSYAGAITGHPDVKYH 240
 DB 175 SLTRGSAVRLVVLALTNFLNHRITPVPPLRGISASGDSPLSYAGAITGHPDVKYH 234
 QY 241 VHEGTEKIMFARERAIISLFGLEAVVLPKESGLGVNGTAVASASWATLSLHDSHMLSLSG 300
 DB 235 VHEGTEKIMFARERAIISLFGLEAVVLPKESGLGVNGTAVASASWATLSLHDSHMLSLSG 294
 QY 301 ALTALTEAMYGQGGSPAPFIHDYCRPHPGQVEYARIRTLSSGSSFAVHEHEEVYKXD 360
 DB 295 SLTAMTVEAMYGHAGSFPFLHDYCRPHPGQVEYARIRTLSSGSSFAVHEHEEVYKXD 354

QY 361 EGI LRDRYPLRTSPQIFGLPELVEDMDMAVSTLSLB-NTTTDNPILDVYENKQTAGGNGFQ 419
 DB 355 EGI LRDRYPLRTSPQIFGLPELVSDLIHAAVLTLEAGSTTDNPLIDVENKTSRGGNFQ 414
 QY 420 ASASISMEKRTSLATLIGKLNFTQCTELLNAAMRGPLSCIAAEDPSLNYHGLDIIH 479
 DB 415 AAIVANTYKTRGLGAQIGKLNFTQCTELLNAAMRGPLSCIAAEDPSLNYHGLDIIH 474
 QY 480 AAYASLGLANPVTTPQPAEMGNQAVNSLALISARTTAPADVLSLLASHLYCTLQA 539
 DB 475 AAYTSELGLANPVTTHQPAEMANOAVNSLALISARTTESNDVLSLLATHLYCVLQA 534
 QY 540 VDLRAMELDFKQFPFLPTLLQOHLG---GLGV-NALALEVKKALKNKLEQTTTDL 595
 DB 535 IDLRALFEFFKQFPGLVSLDQFGSAMTGSULRDELVEKVKTKLAKLEQTNSTDLV 594
 QY 596 PRMHDAFSYATGVVELLSSPSANVTLTAVANKVSAEKALSLTREVNRFPQTPSSQ 655
 DB 595 PRMHDAFSYATGVVELLSSPSANVTLTAVANKVSAEKALSLTREVNRFPQTPSSQ 651
 QY 656 APAHAYLSPRTVLSFVREELGVQARRGVFVGVQCTIGSNRYEAIKGRINHTL 715
 DB 652 SPALSYLSPTQILYAFVREELGVKARRGVFLKQEVTTIGSNVSKYIEAIKGRINHTL 711
 QY 716 VMQLA 720
 DB 712 LKQLA 716

RESULT 9
 AAE27941
 ID AAE27941 standard; protein; 716 AA.
 XX
 AC AAE27941;
 XX
 DT 06-AUG-2003 (revised)
 DT 27-DEC-2002 (first entry)
 XX
 DE Rhodotorula toruloides PAL protein.
 XX
 KM Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
 KM cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
 KM human cytomegalovirus infection; cyclostatic; virucide.
 XX
 OS Rhodosporidium toruloides.
 OS
 PN US2002102712-A1.
 XX
 PD 01-AUG-2002.
 XX
 PF 24-AUG-2001; 2001US-00939408.
 XX
 PR 24-JUL-2000; 2000US-00624693.
 PR 24-JUL-2001; 2001WO-US023270.
 XX
 PA (PCBU-) PCBU SERVICES INC.
 XX
 PI Yoshida RK, Kootstra AB;
 XX
 DR WPI; 2002-690616/74.
 DR N-PSDB; AAD45815.
 XX
 PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
 PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
 PT human immunodeficiency virus or human cytomegalovirus infection.
 XX
 PS Example 3; Page 43-45; 74pp; English.
 XX
 CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC They are useful for treating mammals having diseases, disorders or
 CC conditions that would benefit from treatment with PAL proteins such as

phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or human cytomegalovirus (HCMV) infection. Sequences of the invention are also used in gene therapy. The present sequence is R. cornuoides PAL protein. (Updated on 06-AUG-2003 to correct OS field.)

Sequence 716 AA:

Query Match 72.4%; Score 2614; DB 5; Length 716;

Best Local Similarity 72.0%; Pred. No. 1.7e-219;

Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

1 MAPSLDLSATTTANGFTNGSHAPTKSAAGPTSAIRTPGIDGHAHQSOLEIVQELLSD 60
1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAAVAGSHLPTQTQVDIVYEKMLAA 54
61 PTDVVELSGSYLTVRDVYGAARKGRVVRQNDDEIRAVDKSVDFLKAQCLQNSVYGVTT 120
55 PTDSTLELDGYSLNLDGVASARKGRPVKXDSDEIRKIDKSVDFLSQLSMSYGVTT 114
121 GFGGSADTRTEDAVSLQKALLHQLCGVPTXSXSFSGRGENTLPLEVVRGAMVIRVN 180
115 GFGGSADTRTEDAISLQKALLHQLCGVLPSSPDSFRIGRGLENSLPLEVVRGAMVIRVN 174
181 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGSIASGDLSPISYAGAITGHPDVYKH 240
175 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGSIASGDLSPISYAGAITGHPDVYKH 234
241 VHEGTEKIMFAREALSLFGLAVVLGPKEGGLGVNNGRAVSASMATLSLHSHMLSLSQ 300
235 VHEGTEKIMFAREALSLFGLAVVLGPKEGGLGVNNGRAVSASMATLSLHSHMLSLSQ 294
301 ALTALIVEAMVQCGSFAPFIHDVCRPHPGQVEYARNIRTLISGSSFAVEHEEYKVKD 360
295 SLTAMTVEAMVQCGSFAPFIHDVCRPHPGQVEYARNIRTLISGSSFAVEHEEYKVKD 354
361 EGLRQDRPLRTSPQFLGPIVEDMAYSTLSLE-NTTTDNPILDVYENKQTAHGNFQ 419
355 EGLRQDRPLRTSPQFLGPIVEDMAYSTLSLE-NTTTDNPILDVYENKQTAHGNFQ 414
420 ASAVSISMEKTRIALALIKLNFQCTELLNANNRGLPSCIAEDPSLNYHGKLDIHI 479
415 ASAVSISMEKTRIALALIKLNFQCTELLNANNRGLPSCIAEDPSLNYHGKLDIHI 474
480 AAYASELGLANPVTTTFVQPAEMGQAVNSLALISARRTAENDVSLLSLASHYCTLQA 539
475 AAYASELGLANPVTTTFVQPAEMGQAVNSLALISARRTAENDVSLLSLASHYCTLQA 534
540 VDIRAMELDPKQCFDLPLTLQOHG---TGLDY-NLLALEVKKALKRLEQTTVDLE 595
535 VDIRAMELDPKQCFDLPLTLQOHG---TGLDY-NLLALEVKKALKRLEQTTVDLE 594
596 PRWDAFSYATGIVVELSSPSANVTLTAVANAKVASAEKASISLTRYENRFVQTPSSQ 655
595 PRWDAFSYATGIVVELSSPSANVTLTAVANAKVASAEKASISLTRYENRFVQTPSSQ 651
656 APAAHAYISPTRLVYIFVBEELCVQARRQDVVQVQOETIGSNVRITAEAKDKGINHVL 715
652 SPALSYLSPTQILYAFVBEELGVKARQDVFLGQOEVTTIGSNVSKIVYALKSGRINVL 711
716 VKKLA 720
712 LKKLA 716

```

RESULT 10
AAE20653 standard; protein; 716 AA.

XX AAE20653;
XX AC AAE20653;
XX DT 01-JUL-2002 (first entry)
XX Rhodotorula glutinis phenylalanine ammonia lyase (PAL).

XX Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
XX phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
XX cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
XX LCP; telecommunication; medical device; aerospace application; enzyme;
XX biocatalyst.

OS Rhodotorula glutinis.

PN MO200210407-A1.

XX 07-FEB-2002.

XX 23-JAN-2001; 2001WO-US002099.

XX 27-JUL-2000; 2000US-00627216.

XX (DURO) DU PONT DE NEMOURS & CO E. I.

XX Gatenby AA, Sariastani FS, Tang X, Qi W, Vannelli T;

XX WPI; 2002-280635/32.

XX N-PSDB; AAD33075.

PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT polypeptide, or mutant TAL polypeptide, useful for producing
PT para-hydroxycinnamic acid in recombinant host cell lacking cinnamate
PT hydroxylase.

PS Example 6; Page 92-95; 139pp; English.

XX The invention relates to methods for biological production of para-
XX hydroxycinnamic acid (PHCA). The invention relates to developing of a new
XX biocatalyst for conversion of glucose to PHCA by incorporation of the
XX wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
XX Rhodotorula glutinis into Escherichia coli underlying the ability of the
XX wild-type PAL to convert tyrosine to PHCA. The invention is also directed
XX to developing a new biocatalyst for conversion of glucose to PHCA by
XX incorporation of a new biocatalyst PAL which possesses enhanced tyrosine
XX ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
XX plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
XX Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
XX recombinant techniques. The recombinantly produced PHCA may be used as a
XX monomer for production of liquid crystal polymers (LCP). LCP may be used
XX in electronic connectors and telecommunication and aerospace
XX applications. LCP resistance to sterilizing radiation has also enabled
XX these materials to be used in medical devices as well as chemical, and
XX food packaging applications. The present sequence is Rhodotorula glutinis
XX PAL protein.

XX Sequence 716 AA;

XX Query Match 72.4%; Score 2614; DB 5; Length 716;

XX Best Local Similarity 72.0%; Pred. No. 1.7e-219;

XX Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

1 MAPSLDLSATTTANGFTNGSHAPTKSAAGPTSAIRTPGIDGHAHQSOLEIVQELLSD 60
1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAAVAGSHLPTQTQVDIVYEKMLAA 54
61 PTDVVELSGSYLTVRDVYGAARKGRVVRQNDDEIRAVDKSVDFLKAQCLQNSVYGVTT 120
55 PTDSTLELDGYSLNLDGVASARKGRPVKXDSDEIRKIDKSVDFLSQLSMSYGVTT 114
121 GFGGSADTRTEDAVSLQKALLHQLCGVPTXSXSFSGRGENTLPLEVVRGAMVIRVN 180
115 GFGGSADTRTEDAISLQKALLHQLCGVLPSSPDSFRIGRGLENSLPLEVVRGAMVIRVN 174
181 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGSIASGDLSPISYAGAITGHPDVYKH 240
175 SLTRGSAVRVLVLEALTNFLNHRITPIVPLRGSIASGDLSPISYAGAITGHPDVYKH 234
241 VHEGTEKIMFAREALSLFGLAVVLGPKEGGLGVNNGRAVSASMATLSLHSHMLSLSQ 300

```

Db 235 VVHGKEKILYAREMAMLFNLEPVLGPKGLGVNGTAVASAMTLALHDAHMSLSLQ 294
 Qy 301 ALTALTVEAMVGOOGSFAPIHDVCRPHQOVEVARNIRTLSSGSSFAVEHEEVKXD 360
 Db 295 SLTAMTVEAMVGHAGSFHPIHDVTRPHPIQIEVAGNIRKLEGSRAVHHEEVKXD 354
 Qy 361 EGIIRQDRYPLRTSPQPLGPIVEDMMAVSTLSL-NTTTDNPPLDVENKOTAGGNFQ 419
 Db 355 EGIIRQDRYPLRTSPQPLGPIVEDMMAVSTLSL-NTTTDNPPLDVENKOTAGGNFQ 414
 Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNAAAMNGLPSCLAEDPISLNYHGKGLDHI 479
 Db 415 AAAVANMEXKTRILAGIKLNFTQCTELNAAAMNGLPSCLAEDPISLNYHGKGLDHI 474
 Qy 480 AAAYSEIGHLANPPTTVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTTQA 539
 Db 475 AAAYSEIGHLANPPTTVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTTQA 534
 Qy 540 VDLRAMELDFKKQDPFLPTLLOOHLG---TGLDV-NALALEYKALNKRLEQTTTDE 595
 Db 535 IDLRALIEFEFKQGPALIVSLIOHFGSANTGSNLDDELVEKNTLAKLEQTNISYDLV 594
 Qy 596 PRMHDAFSYATGVVEILLSSPSANVTTLAVNAKVAASAELSLTREVRNRFWQTPSSQ 655
 Db 595 PRMHDAFSYATGVVEILLSSPSANVTTLAVNAKVAASAELSLTREVRNRFWQTPSSQ 651
 Qy 656 APAAVYSPRTVLYSVREELGVQARBGDPVQVQOETIGSNVSRITVIAIKDGINHYL 715
 Db 652 SPALSTYSPRTQILYAVREELGVQARBGDPVQVQOETIGSNVSRITVIAIKDGINHYL 711
 Qy 716 VKMLA 720
 Db 712 LKMLA 716

RESULT 11

AAE16389 ID AAE16389 standard; protein; 716 AA.

AAE16389;

03-APR-2002 (first entry)

Rhodotorula glutinis wildtype phenylalanine ammonia-lyase (PAL).

KM Phenylalanine ammonia-lyase; PAL; enzyme; cinamate; food packing; PHCA;
 KM para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
 KM Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
 KM aerospace application; chemical; red yeast.

Rhodotorula glutinis.

W0200111071-A2.

15-FEB-2001.

03-AUG-2000; 2000MO-US021156.

06-AUG-1999; 99US-0147713P.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Tang X, Vannelli TM, Qi WH, Sarislan S, Gatenby AA;

WPI; 2002-121549/16.

N-PSTDB; AAD26923.

Producing para-hydroxycinnamic acid for producing liquid crystal polymer;
 PT comprises converting cinamate to PHCA, glucose to PHCA by phenyl ammonia
 PT-lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
 activity.

PS Claim 8, Page 65-67; 75pp; English.
 XX The invention relates to a method for producing para-hydroxycinnamic acid
 CC (PHCA). The method comprises converting cinamate to PHCA by converting
 CC glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
 CC route or by generating a new biocatalyst possessing enhanced tyrosine
 CC ammonia-lyase (TAL) activity. The method is useful for the biological
 CC production of PHCA which is useful as a monomer for the production of
 CC Liquid Crystal Polymers (LCP), where the LCP is used in electronic
 CC connectors, telecommunications and aerospace applications. LCP is also
 CC useful in medical devices, as well as chemical and food packing
 CC applications due to its resistance to sterilising radiation. The present
 CC sequence is red yeast, Rhodotorula glutinis (Rhodosporidium toruloides)
 CC wildtype PAL
 CC
 SQ Sequence 716 AA;

Query Match 72.4%; Score 2614; DB 5; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

Qy 1 MAPSIDSLATTLANGFTNGSHAPRTSAAGPTSAIRTPGIDGHAHQOQLTVOELBLSD 60
 Db 1 MAPSIDSLATTLANGFTNGSHAPRTSAAGPTSAIRTPGIDGHAHQOQLTVOELBLSD 54
 Qy 61 PTDVVEELSGYSLITRDVYGAARKGRVVRQNDIEIARVDKSYDELKAOLQNSVGVTT 120
 Db 55 PTDSTLELDGYSNLGADVSAARKGRPVVRQNDIEIARVDKSYDELKAOLQNSVGVTT 114
 Qy 121 GFSGADRTEDAVSLQAL:EHQLCGVTPYSXSFVVGLENTLLEVVVRGMYRVN 180
 Db 115 GFSGADRTEDAVSLQAL:EHQLCGVTPYSXSFVVGLENTLLEVVVRGMYRVN 174
 Qy 181 SLTRGHSAYRLVLEALTNFINRITPIVPLRGSISASGDLSPITYAGAITHGPVYKH 240
 Db 175 SLTRGHSAYRLVLEALTNFINRITPIVPLRGSISASGDLSPITYAGAITHGPVYKH 244
 Qy 241 VHEGTEKIMEAREALISFGLAEVVLGPKGLGVNGTAVASAMTSLHDSHMSLSLQ 300
 Db 235 VHEGTEKIMEAREALISFGLAEVVLGPKGLGVNGTAVASAMTSLHDSHMSLSLQ 294
 Qy 301 ALTALTVEAMVGOOGSFAPIHDVCRPHQOVEVARNIRTLSSGSSFAVEHEEVKXD 360
 Db 295 SLTAMTVEAMVGHAGSFHPIHDVTRPHPIQIEVAGNIRKLEGSRAVHHEEVKXD 354
 Qy 361 EGIIRQDRYPLRTSPQPLGPIVEDMMAVSTLSL-NTTTDNPPLDVENKOTAGGNFQ 419
 Db 355 EGIIRQDRYPLRTSPQPLGPIVEDMMAVSTLSL-NTTTDNPPLDVENKOTAGGNFQ 414
 Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNAAAMNGLPSCLAEDPISLNYHGKGLDHI 479
 Db 415 AAAVANMEXKTRILAGIKLNFTQCTELNAAAMNGLPSCLAEDPISLNYHGKGLDHI 474
 Qy 480 AAAYSEIGHLANPPTTVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTTQA 539
 Db 475 AAAYSEIGHLANPPTTVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTTQA 534
 Qy 540 VDLRAMELDFKKQDPFLPTLLOOHLG---TGLDV-NALALEYKALNKRLEQTTTDE 595
 Db 535 IDLRALIEFEFKQGPALIVSLIOHFGSANTGSNLDDELVEKNTLAKLEQTNISYDLV 594
 Qy 596 PRMHDAFSYATGVVEILLSSPSANVTTLAVNAKVAASAELSLTREVRNRFWQTPSSQ 655
 Db 595 PRMHDAFSYATGVVEILLSSPSANVTTLAVNAKVAASAELSLTREVRNRFWQTPSSQ 651
 Qy 656 APAAVYSPRTVLYSVREELGVQARBGDPVQVQOETIGSNVSRITVIAIKDGINHYL 715
 Db 652 SPALSTYSPRTQILYAVREELGVQARBGDPVQVQOETIGSNVSRITVIAIKDGINHYL 711
 Qy 716 VKMLA 720
 Db 712 LKMLA 716

RESULT 12
 ID ABG71948 standard; protein; 716 AA.
 AC ABG71948;
 DT 27-JAN-2003 (first entry)
 DE R. glutinis wild-type Phenylalanine ammonia-lyase, PAL.
 XX
 XX Enzyme: phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
 KM para-hydroxycinnamic acid; tyrosine ammonia lyase; TAL; tyrosine;
 XX cinnamic acid; PHCA; liquid crystal polymer.
 XX
 OS Rhodotorula glutinis.
 XX
 PN MO200290523-A2.
 XX
 PD 14-NOV-2002.
 PF 03-MAY-2002; 2002MO-US018551.
 XX
 PR 04-MAY-2001; 2001US-0288701P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Qi WM, Sarislaani FS, Tang X;
 XX
 DR WPI; 2003-058928/05.
 XX
 DR N-PSDB; ABS56530.
 XX
 PT Novel recombinant host for producing para-hydroxycinnamic acid and
 PT tyrosine, comprises gene encoding tyrosine ammonia lyase activity and
 XX gene encoding phenylalanine hydroxylase activity.
 XX
 PS Claim 5; Page 48-50; 69pp; English.
 XX
 CC The invention relates to a recombinant host comprising at least one gene
 CC encoding a tyrosine ammonia lyase (TAL) activity, and at least one gene
 CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
 CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
 CC providing the recombinant host; and (2) growing the recombinant organism
 CC in the presence of a fermentable carbon substrate and thus producing para
 CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
 CC providing a recombinant organism comprising at least one gene encoding a
 CC phenylalanine hydroxylase activity; and (2) growing the recombinant
 CC organism in the presence of a fermentable carbon substrate to produce
 CC tyrosine. The recombinant host cell is used for producing para-
 CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
 CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
 CC monomer for the production of liquid crystal polymer. The present
 CC sequence is R. glutinis wild-type Phenylalanine ammonia-lyase, PAL, an
 CC enzyme which catalyses the conversion of phenylalanine into cinnamic acid
 XX
 SQ Sequence 716 AA;
 Query Match 72.4%; Score 2614; DB 6; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY 181 SLTRGSAVRLVLEALTNFLNRIPIVPLRGSIASAGDLSLSYIAGAITGHPDVKA 240
 DB 175 SLTRGSAVRLVLEALTNFLNRIPIVPLRGSIASAGDLSLSYIAAISGHPDKVH 234
 QY 241 VLBGTCKIMFARBAISLFGLEAVVLGPKKGLG;VNGTAVASAMATLSLHDSMLSLSQ 300
 DB 235 VVHEGKEKILYARBAAMLFLNLEPVILGPKKGLG;VNGTAVASAMATLALHADMLSLSQ 294
 QY 301 ALTALTAVAMGQGSFAPFTHDVCRRHPGQVEYARIRITLSSGSPFAVHEBEVKVD 360
 DB 295 SLTAVTAVAMGHSFHPFLHDVTRPHPTQIEVAGIRKLLGSRRAVHHEBEVKXD 354
 QY 361 EGLIKQDRYPRRTSPQFVGLVEDMKAYSTLSLE-NNITTDNPLDVENKQTARAGNFQ 419
 DB 355 EGLIKQDRYPRRTSPQFVGLVEDMKAYSTLSLE-NNITTDNPLDVENKQTHHGNFQ 414
 QY 420 ASAVSISEKTRRLALIGKINFTQCTELNMAANRGLPSCLAEDPSLHYHGKLDIHI 479
 DB 415 AAAVANTMEKTRLGIAQIGKINFTQCTELNMAANRGLPSCLAEDPSLHYHGKLDIHA 474
 QY 480 AAYASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTTQA 539
 DB 475 AAYTSELGLANPVTTHQPAEMANQAVNSLALISARRTESNDVLSLLATHLYCVLQA 534
 QY 540 VDLRAVELDPKKQDPPLPTLLOQHIG---TGLDV-NALALEYKKAANKLEQTTVDLE 595
 DB 535 IDLRALIEFEFKQFGPALVSLIDHFGSAMTGNLRELVEKNKTLAKRLEQNTSDLV 594
 QY 596 PRWDAFSGVAGTVELLSSPSANVTITAVAMKVASAKAISLTREVNRRFMOPTPSQ 655
 DB 595 PRWDAFSGVAGTVEVLSST---SLSLAANAMKVAASGAILSTHQVETFPASASTS 651
 QY 656 APAAHYLSPTTRVLYSFVREELGVQARRGDVFVGVOETISNSRITYEAIKGRINHYL 715
 DB 652 SPALSTLSPTQILYAFVREELGVYARRGDVFLGKQEVLTIGSNVSKITYEAIKGRINNVL 711
 QY 716 VKMLA 720
 DB 712 VKMLA 716
 RESULT 13
 ADE36054
 ID ADE36054 standard; protein; 716 AA.
 AC ADE36054;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rhodosporidium glutinis PAL protein SEQ ID NO:2.
 XX
 XX recombination; nucleic acid template; recombinant polypeptide;
 KM altered property; combined gene.
 XX
 OS Rhodosporidium.
 XX
 PN WO2003072743-A2.
 XX
 PD 04-SEP-2003.
 XX
 PF 26-FEB-2003; 2003MO-US005708.
 XX
 PR 26-FEB-2002; 2002US-0360279P.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Milano J, Tang X;
 XX
 DR WPI; 2003-721766/68.
 XX
 DR N-PSDB; ADE36053.
 PT Recombination of nucleic acid templates, useful e.g. for producing
 PT enzymes with altered properties, is based on cyclic extension of unpaired

Primer.
 Dislosure; SEQ ID NO 2; 481bp; English.
 The present invention describes a method for the recombination of nucleic acid (NA) templates. The method comprises: (a) providing at least two double-stranded templates (T1, T2) having different 5' and 3' regions in their sense strands; (b) contacting the templates with at least one each of forward and reverse primers that anneal, respectively, only to the 3'-region of the antisense strand of T1 and to the 3'-region of the sense strand of T2; (c) extending the primers by no more than 1000 nucleotides (nt); (d) separating extended primers from their templates; (e) reannealing the extended primers to either template; and (f) repeating cycles of steps (b)-(e) until at least one full-length extension product, i.e. a recombination of T1 and T2, is obtained. Also described: (1) a method starting from at least one each of antisense and sense single-stranded templates; (2) generating a recombinant polypeptide (I) with altered properties by expressing recombination products of the new processes and screening the polypeptides formed for properties different from those of polypeptides from either of the templates; and (3) (1) produced by method (2). The methods are useful for the recombination of nucleic acid templates of interest, and generation of a recombinant polypeptide having altered properties. The method can be used to make combined genes that express proteins e.g. enzymes, cytokines, growth factors, viral proteins or microbial antigens with altered properties, e.g. stability, activity or specificity. The present sequence is used in the exemplification of the present invention.

Sequence 716 AA;
 Query Match 72.4%; Score 2614; DB 7; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1.7e-219;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

1 MABSLDLSLTTANGFTNGSHAPTKSAAPTSAURRTGDLGHAHOSOLEIYOELISD 60
 1 MAESLDSISHSFANG-----VASAKQAVNGASTVLAAGSHLPTQYTDYIEKMLAA 54
 61 PTDVVELSGYSTLVNDVVGAKRGRRVQVNDDEIRARVKSVDLFAQIQNSYVGT 120
 55 PTDSTLHLDGYSINLNDVVASAKRGRRVQVNDDEIRARVKSVDLFAQIQNSYVGT 114
 121 GFGGSADTRTEDAVSLQKALLIHOLOGVPTTSXSFVSRCENTLPLEVVRGAMVRVN 180
 115 GFGGSADTRTEDAISIQKALLHQLCGVLPSSFSDFRGRGELNSLPLEVVRGAMVRVN 174
 181 SLRGSASAVRLVYLALTNFLNHRITPIVPLKGSISASGDLSPISYAGAITGHPDVKA 240
 175 SLRGSASAVRLVYLALTNFLNHRITPIVPLKGSISASGDLSPISYAGAITGHPDVKA 234
 241 VLHGETEKIMFAREASLFGLEAVVLGPKEGGLVNGTAVASAMATSLHDSHMLSLSSQ 300
 235 VVHGETEKILVAREANALFNLEPVLDGREGIGLVNGTAVASAMATSLHDSHMLSLSSQ 294
 301 ALTALTVEMVQGGSPAFHIDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKVKXD 360
 295 SLTAMTVEMVGHAGSFPHFDVTRPHPTQIEVAGNIRKLKESRFAVHHEEVKVKXD 354
 361 EGIHQDRYPLRTSPQIFGIVEDMMHAYSTLSLE-NTTNDNLLDVENKQDRAHGNFQ 419
 355 EGIHQDRYPLRTSPQIFGIVEDMMHAYSTLSLE-NTTNDNLLDVENKQDRAHGNFQ 414
 420 AASAVISMEKTRIALALIGKINFTQCTELNNAANRGLPSCIAEDPSLNYHGKGLDIH 479
 415 AASAVISMEKTRIALALIGKINFTQCTELNNAANRGLPSCIAEDPSLNYHGKGLDIH 474
 480 AAYASISLIGLANPVTTPVOPAEAGNQAANISALISARTANANDVLSLLASHIXCTLQA 539
 475 AAYTSLIGLANPVTTPVOPAEAGNQAANISALISARTANANDVLSLLASHIXCTLQA 534
 540 VDLRAMELDFKKQDPDLPTLLQOHTG---TGLDV-NALALEVKAALKRLEQTTDYDE 595
 535 IDLRAITEFEFKQDFEPALVSLIDHFGSAGMGSNLRELIVKWKTLAKGLQETNSVDLV 594

596 PRWHDAPSVATGTVNVELLSSSSPSANVTLTVANAKVYAGAEKALISTREVRNRFQCTPSQ 655
 595 PRWHDAPSVATGTVNVELLSSSSPSANVTLTVANAKVYAGAEKALISTREVRNRFQCTPSQ 651
 656 APAAVILSPRTVLYSFVARELGVARQDVGVQCEITGSNVSRIYEALIKDGRINHL 715
 652 SPALSLSTRTOILVAFVARELGVARQDVGVQCEITGSNVSRIYEALIKDGRINHL 711
 716 VKMLA 720
 712 LKMLA 716

RESULT 14
 AAE20660
 ID AAE20660 standard; protein; 716 AA.
 AC AAE20660;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Rhodotorula glutinis mutant tyrosine ammonia lyase (TAL), RM120-2 #1.
 XX
 KM Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KM phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KM cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
 KM LCP; telecommunication; medical device; aerospace application; enzyme;
 KM biocatalyst; mutant; variant; mutein.
 XX
 OS Rhodotorula glutinis.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 198
 FT /label= Asp, Asn, Glu, Gln
 FT Misc-difference 540
 FT /label= Thr, Ala, Ser, Pro, Gly
 PN WO200210407-A1.
 PD 07-FEB-2002.
 PP 23-JAN-2001; 2001WO-US002099.
 PR 27-JUL-2000; 2000US-00627216.
 PA (DUPO) DU POINT DE NEMOURS & CO E. I.
 PI Gatenby AA, Saritasani FS, Tang X, Qi W, Vannelli T;
 XX WPI; 2002-280635/32.
 XX Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT polypeptide, or mutant TAL polypeptide, useful for producing
 PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
 PT hydroxylase.
 PS Claim 2, Page 126-128, 139pp; English.
 PS The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
 CC plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
 CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
 CC recombinant techniques. The recombinantly produced PHCA may be used as a
 CC monomer for production of liquid crystal polymers (LCP). LCP may be used

CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC mutant TAL

XX Sequence 716 AA;

Query Match 72.4%; Score 2613; DB 5; Length 716;

Best Local Similarity 72.0%; Pred. No. 2.1e-219; Mismatches 522; Conservative 85; Matches 104; Indels 14; Gaps 5;

1 MAFSLDLSATTTANGFTNGSHAAPTKSAGPTSAIRRTGGDGAHQSOEIQELISD 60
 1 MAFSLDLSISHPANG-----VASAKQAVNGASTMLAAVAGSLPTTQVTVQVDIYKMLAA 54
 61 PTDVVELSGYSTLRDVVGAARKGRVAVQNDDEIRARVDKSVDFLKAQONSYYGVT 120
 55 PTDSTLELDGYSINLDGVSAARKGRPVVKDSDERIKSDVSEFLRSQLSMSYGVTT 114
 121 GFGSADRTEDAVSLQKALIEHQLCGVPTXSXSFSGRGLENTLPLEVVRGAMVIRVN 180
 115 GFGSADRTEDAVSLQKALIEHQLCGVLPSSFDFRUGLENSLPLEVVRGAMVIRVN 174
 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDISPLSYAGAITGHPDVKH 240
 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDISPLSYAGAITGHPDVKH 234
 241 VMEGTEKIMFARPAEASLFGLENVYIGPREGIGLVNCTRVASMTLSLHSHMLSLISQ 300
 235 VVEGGEKILYAEAAALFNLEVVYGPREGIGLVNCTRVASMTLSLHSHMLSLISQ 294
 301 ALTALTEAMVWGQGSFAPFIHDCRPHGOVEVARNIRTTLSGSSFAVEHEEYKXVD 360
 295 SLTAMTEAMVWAGSFPFPHLDVTRPHPTQIEVAGNIRKLESGRPAVHHEEYKXVD 354
 361 EGIIRDPRPLRSPFLGPIYEDMMHANSSTLE--NNTTNPILLDENKOTAGCNFQ 419
 355 EGIIRDPRPLRSPFLGPIYEDMMHANSSTLE--NNTTNPILLDENKOTAGCNFQ 414
 420 ASAVSISMEKTRTALIGKLNTOCTELLNAAHNGRLPSCLAADPSLHNGKLDIHI 479
 415 AAANVTMETRIGLQIGKLNTOCTELLNAAHNGRLPSCLAADPSLHNGKLDIHI 474
 480 AAVASELGHIANVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHIYCTLOA 539
 475 AATSEIGHIANVTTHVQPAEMANOAVNSLALISARRTAEANDVLSLLASHIYCTLOA 534
 540 VDIRAMELDPKQFDBLLPTLLQOHG---TGLDV--NALALEVKALKRLEQTTYNLE 595
 535 IDRAAEFEKQFGPAIVSLIDQHFQAMTGSNIRDELVEKVNKTLLAKRLEQTNISYDLV 594
 596 PRMHDAFSYATGVVBLSSSPSANVTTLTAVANAKVASAEKAISLTREVRNRFQTPSSQ 655
 595 PRMHDAFSYATGVVBLSSSPSANVTTLTAVANAKVASAEKAISLTREVRNRFQTPSSQ 651
 656 APHAHVLSPETRVLYSFPVEELGVQARRDVGVQOQETIGSNVRITPAIKGFINHVL 715
 652 SPALSYLSPETRVLYSFPVEELGVQARRDVGVQOQETIGSNVRITPAIKGFINHVL 711
 716 VKMLA 720
 712 LKMLA 716

RESULT 15

ABG71951 standard; protein; 716 AA.

XX ABG71951,
 XX 27-JAN-2003 (first entry)
 XX

DE R. glutinis PAL/TAL mutant #3.

XX Enzyme; phenylalanine ammonia-lyase, PAL; phenylalanine hydroxylase, PAH;
 XX para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
 XX tyrosine; cinnamic acid; PHCA; liquid crystal polymer; mutant; mutein.

XX Rhodotorula glutinis.

XX Synthetic.

XX Key Location/Qualifiers
 XX MISC-difference 198
 XX MISC-difference 540
 XX MISC-difference 540
 XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

XX MISC-difference 540

QY 241 VHEGTEKIMFARBAISLPGLEAVVLGPKEGLVNGTAVSASMATLSJHDSHMLSLLSQ 300
 Db 235 VHEGTEKIMFARBAISLPGLEAVVLGPKEGLVNGTAVSASMATLSJHDSHMLSLLSQ 294
 QY 301 ALTAITVEAMVGOQGSFAPFIHDVCEPHQGVAVANNIRTLGSGSFAYEHEEYKXD 360
 Db 295 SLTMTVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGSFAYHHEEYKXD 354
 QY 361 EGIIRORRYPLRTISPOFLGVLVEDMHAISTLSLE-NNTTNDPLIDVENKQTAHGNFQ 419
 Db 355 EGIIRORRYPLRTISPOFLGVLVEDMHAISTLSLE-NNTTNDPLIDVENKQTAHGNFQ 414
 QY 420 ASAVSISMEKTRIALALIGKLNFTOCTELINAMNGLPSCLAEDPSSLNYHKGJLDIHI 479
 Db 415 AAAYANTWEXTRJGLAIGKLNFTOCTELINAMNGLPSCLAEDPSSLNYHKGJLDIHI 474
 QY 480 AAAYSELGHLANPVTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHIYCTLOA 539
 Db 475 AAAYSELGHLANPVTFVQPAEMGNQAVNSLALISARTAEANDVLSLLASHIYCTLOA 534
 QY 540 VDLZAMELDFKKQDPLLTLLQOHIG--TGUDV-NALALEYKALNKRELEQTTIDLE 595
 Db 535 IDLZAMELDFKKQDPLLTLLQOHIG--TGUDV-NALALEYKALNKRELEQTTIDLE 594
 QY 596 PRWHDATFATGTVBELSSSPSANTVLTAVNAMKYASAEKAIISLTREYANRFMOQPSQ 655
 Db 595 PRWHDATFATGTVBELSSSPSANTVLTAVNAMKYASAEKAIISLTREYANRFMOQPSQ 651
 QY 656 APAAAYLSPTRTVLYSFVREBELGVQARQGVFVQOQETIGSNVSRITYEALIKGRINHVL 715
 Db 652 SPALSYLSPTRTVLYSFVREBELGVQARQGVFVQOQETIGSNVSRITYEALIKGRINHVL 711
 QY 716 VKMLA 720
 Db 712 LKMLA 716

Search completed: September 9, 2004, 10:04:00
 Job time : 92.6307 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:01:03 ; Search time 23.4025 Seconds
(without alignments)
1588.322 Million cell updates/sec

Title: US-09-939-408a-13

Sequence: 1 MAPSLDLSLTIANGFTNGS.....RIEYAKGRINHYLVMLA 720

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCUTUS.COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3608	99.9	720	4 US-09-624-693A-13	Sequence 13, Appl
2	2770	76.7	726	4 US-09-624-693A-21	Sequence 21, Appl
3	2614	72.4	716	4 US-09-624-693A-19	Sequence 19, Appl
4	2614	72.4	716	4 US-09-627-216A-8	Sequence 8, Appl
5	2614	72.4	716	4 US-09-765-873A-8	Sequence 8, Appl
6	2613	72.4	716	4 US-09-765-873A-35	Sequence 35, Appl
7	2612	72.4	716	4 US-09-627-216A-10	Sequence 10, Appl
8	2612	72.4	716	4 US-09-765-873A-38	Sequence 38, Appl
9	2607	72.2	716	4 US-09-765-873A-36	Sequence 36, Appl
10	2603	72.1	716	4 US-09-765-873A-37	Sequence 37, Appl
11	2603	72.1	716	4 US-09-765-873A-34	Sequence 34, Appl
12	2599	72.0	716	4 US-09-765-873A-33	Sequence 33, Appl
13	2579	71.3	716	4 US-09-765-873A-32	Sequence 32, Appl
14	2573	71.3	713	4 US-09-624-693A-17	Sequence 17, Appl
15	2499.5	69.2	713	4 US-09-624-693A-15	Sequence 15, Appl
16	1346.5	37.3	740	4 US-09-615-192A-337	Sequence 328, App
17	716	19.8	424	4 US-09-540-236-285	Sequence 285, App
18	524	14.5	525	4 US-09-328-352-509	Sequence 509, App
19	508.5	14.1	515	4 US-09-489-039A-13575	Sequence 13575, A
20	503.5	13.9	508	4 US-09-282-991A-33046	Sequence 33046, A
21	500.5	13.9	515	4 US-09-282-991A-33046	Sequence 33046, A
22	486	13.5	412	4 US-09-615-192A-330	Sequence 330, App
23	397	11.0	520	4 US-09-252-991A-33049	Sequence 33049, A
24	241	6.7	226	4 US-09-615-192A-337	Sequence 327, App
25	211.5	5.9	97	4 US-09-615-192A-329	Sequence 329, App
26	129	3.6	2123	3 US-08-968-685A-10	Sequence 10, Appl
27	127	3.5	2586	3 US-08-936-135-2	Sequence 2, Appl

28	124	3.4	857	4 US-09-252-991A-23956	Sequence 23956, A
29	118.5	3.3	1832	3 US-09-335-409-4	Sequence 4, Appl
30	118.5	3.3	1832	4 US-09-568-102-4	Sequence 4, Appl
31	118.5	3.3	1832	4 US-09-567-969-4	Sequence 4, Appl
32	118.5	3.3	1832	4 US-09-568-480-4	Sequence 4, Appl
33	118.5	3.3	1832	4 US-09-568-486-4	Sequence 4, Appl
34	118.5	3.3	1832	4 US-09-568-472-4	Sequence 4, Appl
35	118.5	3.3	1832	4 US-09-567-899-4	Sequence 53, Appl
36	118	3.3	647	4 US-09-613-303-53	Sequence 331, Appl
37	118	3.3	647	4 US-10-267-311-53	Sequence 331, Appl
38	117	3.2	132	4 US-09-615-192A-331	Sequence 4, Appl
39	117	3.2	1839	2 US-09-172-977-4	Sequence 4, Appl
40	117	3.2	1839	4 US-09-404-108-4	Sequence 4, Appl
41	116.5	3.2	1239	1 US-08-026-1385-3	Sequence 498, App
42	116.5	3.2	1589	4 US-09-543-681A-4998	Sequence 30639, A
43	115.5	3.2	366	4 US-09-252-991A-30639	Sequence 54, Appl
44	115.5	3.2	1214	2 US-08-231-193A-54	Sequence 54, Appl
45	115.5	3.2	1214	2 US-08-466-273A-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1
US-09-624-693A-13
Sequence 13, Application US/09624693A
Patent No. 6355468
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
FILE REFERENCE: 294/9/500NSC
CURRENT APPLICATION NUMBER: US/09/624,693A
CURRENT FILING DATE: 2000-07-24
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodococcus graminis
US-09-624-693A-13

Query Match	99.9%	Score 3608	DB 4	Length 720
Best Local Similarity	99.9%	Pred. No. 0		
Matches 719	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	1	MAPSLDLSLTIANGFTNGSHAAPTKSAAGPTSAIRTPGLDGHAAHQSOLFIVEELSD	60	
DB	1	MAPSLDLSLTIANGFTNGSHAAPTKSAAGPTSAIRTPGLDGHAAHQSOLFIVEELSD	60	
QY	61	PTDDVVELSGSYLVYADVGAARKGRVAVQNDDEIRARVDKSVDFLKAQOLNSVYGYTT	120	
DB	61	PTDDVVELSGSYLVYADVGAARKGRVAVQNDDEIRARVDKSVDFLKAQOLNSVYGYTT	120	
QY	121	GFGSADRTEDAVSLQKALIEHOLCGVPTISXSSFSYVGRLNTLPPEVNGAVYIRVN	180	
DB	121	GFGSADRTEDAVSLQKALIEHOLCGVPTISXSSFSYVGRLNTLPPEVNGAVYIRVN	180	
QY	181	SLTRGSAVRLVLEALTNPLNHRITPIYPLFGSTISASDLSPLSYTAGATGHDVYKH	240	
DB	181	SLTRGSAVRLVLEALTNPLNHRITPIYPLFGSTISASDLSPLSYTAGATGHDVYKH	240	
QY	241	VHEGTEKIMFAEATSLPGLAVVLGPKEGLVNGVTSASMTLSLHDSHMLSLLSQ	300	
DB	241	VHEGTEKIMFAEATSLPGLAVVLGPKEGLVNGVTSASMTLSLHDSHMLSLLSQ	300	
QY	301	ALTALIVEAMVGGQGFAPFIHDVCRPHGQVAVARNITLLSGSFAYEHEEVKYKD	360	
DB	301	ALTALIVEAMVGGQGFAPFIHDVCRPHGQVAVARNITLLSGSFAYEHEEVKYKD	360	
QY	361	EGILRQRYLRFSPOFLGIVEDMNAVSTLSLNNITTDNPLDVENKQAHGNGFQA	420	

Db 361 EGIIRQDRYPLRTSPQPLGVLVEDMMAVSTLSLENNNTTNDNLDLVEKQTAHAGNFDA 420
Qy 421 SAVSISMEKRLALALIGKLNFTQCTELLNMAANRGPSGLAEDPSLHYHKGJDIH 480
Db 421 SAVSISMEKRLALALIGKLNFTQCTELLNMAANRGPSGLAEDPSLHYHKGJDIH 480
Qy 481 AYAASELGHLANPVTTFVQPAEMGQAVNSIALISARTEABANDVLSLLASHLYCTLOAV 540
Db 481 AYAASELGHLANPVTTFVQPAEMGQAVNSIALISARTEABANDVLSLLASHLYCTLOAV 540
Qy 541 DLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTDLDEPRWD 600
Db 541 DLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTDLDEPRWD 600
Qy 601 AFSTATGTVEELLSSPSSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSSQAPAA 660
Db 601 AFSTATGTVEELLSSPSSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSSQAPAA 660
Qy 661 YLSPTRVLSFVREELGVQARRGDVFGVQOCTIGSNVSRLEYAIIKGRINHTLVKLA 720
Db 661 YLSPTRVLSFVREELGVQARRGDVFGVQOCTIGSNVSRLEYAIIKGRINHTLVKLA 720

RESULT 2

US-09-624-693A-21
; Sequence 21, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocotra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; NAME/KEY: SITE
; LOCATION: (12, 16, 17, 20, 24, 25, 28 - 36, 38, 42, 47, 48, 56, 57, 62, 66,
; OTHER INFORMATION: "Xaa" means any amino acid; "Xaa" means no consensus at that
; OTHER INFORMATION: Position
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19
US-09-624-693A-21

Query Match 76.7%; Score 2770; DB 4; Length 726;
Best Local Similarity 77.7%; Pred. No. 1.3e-262;

Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

Qy 1 MAPSLDSLATTLANGTNGSHAAPTAKAAPTSAIRRTPTGLDGAHAHQSLVEQLSLD 60
Db 1 MAPSLDSLATTLANGTNGSHAAPTAKAAPTSAIRRTPTGLDGAHAHQSLVEQLSLD 60
Qy 61 P-TDDVVELSGYSLTVRDVGAARKGRVRYQNDDEIRARVDKSVDFLXQLONSVYGV 115
Db 61 P-TDDVVELSGYSLTVRDVGAARKGRVRYQNDDEIRARVDKSVDFLXQLONSVYGV 115
Qy 120 TGGGSAADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVGAMVIRV 179
Db 120 TGGGSAADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVGAMVIRV 179
Qy 121 TGGGSAADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVGAMVIRV 180
Db 121 TGGGSAADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVGAMVIRV 180
Qy 180 NSITRGHSAVRLVNLVFNFLNRTIPVPLRGSISASGDLSPSLVYIAGATGHPVYK 239
Db 180 NSITRGHSAVRLVNLVFNFLNRTIPVPLRGSISASGDLSPSLVYIAGATGHPVYK 239
Qy 240 HVLHGTGEXIMFARBAISLFGLEAVVLGPKEGGLVNGTAVASAMATLSLHDSHMLSL 299
Db 240 HVLHGTGEXIMFARBAISLFGLEAVVLGPKEGGLVNGTAVASAMATLSLHDSHMLSL 299

Db 241 HVHHEKEXIMFARBAISLFGLEAVVLGPKEGGLVNGTAVASAMATLSLHDSHMLSL 300
Qy 300 QALTALTVEMAMGQOQSFPIHDYCRPHPGQVEVARIIRTLSSSSFAVHEEVEYK 359
Db 301 QALTALTVEMAMGQOQSFPIHDYCRPHPGQVEVARIIRTLSSSSFAVHEEVEYK 360
Qy 360 DEGLIRQDRYPLRTSPQPLGVLVEDMMAVSTLSLENNNTTNDNLDLVEKQTAHAGNF 418
Db 361 DEGLIRQDRYPLRTSPQPLGVLVEDMMAVSTLSLENNNTTNDNLDLVEKQTAHAGNF 420
Qy 419 QASAVSISMEKRLALALIGKLNFTQCTELLNMAANRGPSGLAEDPSLHYHKGJDIH 478
Db 421 QASAVSISMEKRLALALIGKLNFTQCTELLNMAANRGPSGLAEDPSLHYHKGJDIH 480
Qy 479 IAAVASELGHLANPVTTFVQPAEMGQAVNSIALISARTEABANDVLSLLASHLYCTLO 538
Db 481 IAAVASELGHLANPVTTFVQPAEMGQAVNSIALISARTEABANDVLSLLASHLYCTLO 540
Qy 539 AYLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVNA---LAEVKKALNKRLEQTTTYDL 594
Db 541 AYLRAEMELDFKKQFDPPLPTLLQOHLGTGLDVNA---LAEVKKALNKRLEQTTTYDL 600
Qy 595 EPRMHDASFATGTVEELLSSPSSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSS 654
Db 601 EPRMHDASFATGTVEELLSSPSSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSS 660
Qy 655 QAPAAHYVSPRTVLSFVREELGVQARRGDVFGVQOCTIGSNVSRLEYAIIKGRINHTLV 714
Db 661 QAPAAHYVSPRTVLSFVREELGVQARRGDVFGVQOCTIGSNVSRLEYAIIKGRINHTLV 720
Qy 715 LVKMLA 720
Db 721 LVKMLA 726

RESULT 3

US-09-624-693A-19
; Sequence 19, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kocotra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodococcus toruloides
US-09-624-693A-19

Query Match 72.4%; Score 2614; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 2.6e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

Qy 1 MAPSLDSLATTLANGTNGSHAAPTAKAAPTSAIRRTPTGLDGAHAHQSLVEQLSLD 60
Db 1 MAPSLDSLATTLANGTNGSHAAPTAKAAPTSAIRRTPTGLDGAHAHQSLVEQLSLD 60
Qy 61 P-TDDVVELSGYSLTVRDVGAARKGRVRYQNDDEIRARVDKSVDFLXQLONSVYGV 120
Db 61 P-TDDVVELSGYSLTVRDVGAARKGRVRYQNDDEIRARVDKSVDFLXQLONSVYGV 120
Qy 121 TGGGSAADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVGAMVIRV 180
Db 121 TGGGSAADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVGAMVIRV 180
Qy 180 NSITRGHSAVRLVNLVFNFLNRTIPVPLRGSISASGDLSPSLVYIAGATGHPVYK 239
Db 180 NSITRGHSAVRLVNLVFNFLNRTIPVPLRGSISASGDLSPSLVYIAGATGHPVYK 239
Qy 240 HVLHGTGEXIMFARBAISLFGLEAVVLGPKEGGLVNGTAVASAMATLSLHDSHMLSL 299
Db 240 HVLHGTGEXIMFARBAISLFGLEAVVLGPKEGGLVNGTAVASAMATLSLHDSHMLSL 299

QY	61PNDVDELISGVSLIVEDVVGAAKRGRRVRVONDEIIPARDKSDVDFLKADLOKSVYGVTT	120
Db	55 PFDSTLELDGSLINDGVNVAARKGRVYRKQSDDEISKIDKSEVPIRSLQSLMSYGVTT	114
QY	121 GFGGSADTRTDDAVSLQKALLIEHOLCVPTBTSKSSFSVSGLENTTLPLEVVGAMVIRVN	180
Db	115 GFGGSADTRTDDAISLQKALLIEHOLCVLPSSPFSFPLGGLKLENSLPLEVVGAMTIRVN	174
QY	181 SLTRGHSARLVLYLFAITNFMNHTITPIVPLRGISAGSGLSPFSYTAGIITGHPVKYH	240
Db	175 SLTRGHSARLVLYLFAITNFMNHTITPIVPLRGITISAGSLSPFSYTAIISGHPSKYH	234
QY	241 VLHEGTEKIMFARAAISLIFGLAEAVVLGPKKEGLGLVNTAVASAGMATLSLDSHMLSLGQ	300
Db	235 VVHEKEKILVAREEMALFNLPEVVLGPKKEGLGLVNTAVASAGMATLADAWLSLSLQ	294
QY	301 AITALLTEAMNGQGSFAPFTIHDIYCRPHPGQVEVYAKIRITLLSGSSPYVHEEEVYAKKD	360
Db	295 SLTAMTVEAMNGHAGSFAPFLHDVTRPHPTQIEVAGNIRKLLEGSRFAVHEEEVYAKKD	354
QY	364 EGILRQDRYPARTSPQFGLPVEDMMAVSTLSLE-NNTTTNDPLDVENKQTAHAGSNFQ	419
Db	355 EGILRQDRYPARTSPQWIGPLVSDLIHNAVLTIEAQSSTTDNPLDVENKTHHGSNFQ	414
QY	420 ASAASISMEKTRLALALIGKINFTQCTBELNAAAMRGLPSCLAEDPSLNYHGKLDIHT	479
Db	415 AAASVANTMEKTRLGLAQGKINFTQCTEMNLAGNMRGLPSCLAEDPSLSYHOGGLDIAA	474
QY	480 AAYASELGHIANPVYTFYQPAEMNGQAVNSLATSARTRAEADVLSLLASHYCTLOA	539
Db	475 AAYTSELGHIANPVYTHQPAEMNAQAVNSLATSARTRTESNDVLSLLATHYCYLOA	534
QY	540 VDLRAMELDFKKQDPDLPLPTLLOOHLG---TGLDV-NALALEVKALNKLSLQTTTYDL	595
Db	535 IDLRALIEEFKKQFQPPAVLSLIDHFGSAMGSLRDELVEKYNKTLAKLEQNTSYDLV	594
QY	596 PRMHDAFSAATGTVVELLSSEPSANVTITAYNAKKVSAEALISLTVENNRFPQTSSQ	655
Db	595 PRMHDAFSPAAGTVEVLST---SLSLAAVNAKVAASALSILTQVRETFEWSAAS	651
QY	656 APAAHATLSPRTVLISFVREELGVQARGDVFVVOQETIGSNYSRIYEAIKQGRINHL	715
Db	652 SPALSTLSPRTQILYAFVREELGVQARGDVFILKQVETIGSNYSKIYEAIKSGRIINVL	711
QY	716 VKMLA 720	
Db	712 LKMLA 716	
RESULT 5		
US-09-765-873A-8		
; Sequence 8, Application US/09765873A		
; Patent No. 6521748		
; GENERAL INFORMATION:		
; APPLICANT: Tang, Xiao-Song		
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID		
; FILE REFERENCE: BCI009 US CIP		
; CURRENT APPLICATION NUMBER: US/09/765,873A		
; PRIOR FILING DATE: 2001-01-19		
; PRIOR APPLICATION NUMBER: US 09/627,216		
; PRIOR FILING DATE: 2000-07-27		
; PRIOR APPLICATION NUMBER: US 60/147,719		
; PRIOR FILING DATE: 1998-08-06		
; NUMBER OF SEQ ID NOS: 38		
; SOFTWARE: Microsoft Office 97		
; SEQ ID NO 8		
; LENGTH: 716		
; TYPE: PRT		
; ORGANISM: Rhodocornia glutinis		
US-09-765-873A-8		
Query Match	72.4%;	Score 2614; DB 4; Length 716;
Best local similarity	72.0%;	Pred No. 2,66-247;

```
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
QY 1 MAPSLDLSLATTLANGFTNGSHAAPTKSAAGPTSLRTPGLDGHAAQSLIEVOELSD 60
Db 1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDDVVELSGYSLTVRDVVGAAKRGRRVQVONDEIRARVDKSYDPLKAQLONSYGVTT 120
Db 55 PTDSTLELDGYSLNLDGVVSAARKGRPVKVDSDSEIRSKIDKSVFELRSQLSMSYGVTT 114
QY 121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVRGAMVIRN 180
Db 115 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVRGAMVIRN 174
QY 181 SLTRGSAVRILVLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGAITGHPDVKH 240
Db 175 SLTRGSAVRILVLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGAITGHPDVKH 234
QY 241 VLHGETKIMFARBAISLFGLEAVLIGPKEGGLGVNGTAVASAMTSLSDSHMLSLSQ 300
Db 235 VVHGEKEXILYAREMALEFNLPEVLGPEGLGVNGTAVASAMTSLSDSHMLSLSQ 294
QY 301 ALTALTEAMVGOQGSFAPFIHDVCRPHQOVEVARNIRTLSSGSSFAVEHEEVKXKD 360
Db 295 SLTAMTEAMVGHAGSFPHPLHDVTRPHPTQIEVAGNIRKLLEGRFAVHHEEVKXKD 354
QY 361 EGLIKODRYPLEPRTSPQIFGLPVEDMMAVSTLSLE-NNTTNDNPLDVENKQTRAGSNFQ 419
Db 355 EGLIKODRYPLEPRTSPQIFGLPVEDMMAVSTLSLE-NNTTNDNPLDVENKQTRAGSNFQ 414
QY 420 ASAASISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDHI 479
Db 415 AAAAVANTMEKTRILGIAQIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDHI 474
QY 480 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534
QY 540 VDLRAAMELDPKKQOPPLPPTLLOOHLG---TGLDV-NALALEVKKALKRLEOQTTTYLE 595
Db 535 IDLRAXEFBPKKQOPPAIVSLIDQFSGAMTGSNLRDELVEKYNKTLAKLEQNTSIDLV 594
QY 596 PRMHDAFSFAGTVEVLST---SLSLAAVANAKVAAESAIISLTQVRETFMSAATS 655
Db 595 PRMHDAFSFAGTVEVLST---SLSLAAVANAKVAAESAIISLTQVRETFMSAATS 651
QY 656 APAHAIVSPRTVLYSPFREELGVQARRGDVGVQOETIGSNVSRITYEAIKQGRINHL 715
Db 652 SPALSYLSPTQILYAFVREELGVKARRGDVFLGQEVYIGSNVSKITYEAIKGRINHL 711
QY 716 VKMLA 720
Db 712 LKMLA 716
RESULT 6
US-09-765-873a-35
; Sequence 35, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: ECI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 716
```

```
TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X= Asp, Asn, Glu, Gln
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873a-35
Query Match 72.4%; Score 2613; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 3.2e-247;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;
QY 1 MAPSLDLSLATTLANGFTNGSHAAPTKSAAGPTSLRTPGLDGHAAQSLIEVOELSD 60
Db 1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDDVVELSGYSLTVRDVVGAAKRGRRVQVONDEIRARVDKSYDPLKAQLONSYGVTT 120
Db 55 PTDSTLELDGYSLNLDGVVSAARKGRPVKVDSDSEIRSKIDKSVFELRSQLSMSYGVTT 114
QY 121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVRGAMVIRN 180
Db 115 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSFSFGRLGENTLPLEVVRGAMVIRN 174
QY 181 SLTRGSAVRILVLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGAITGHPDVKH 240
Db 175 SLTRGSAVRILVLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGAITGHPDVKH 234
QY 241 VLHGETKIMFARBAISLFGLEAVLIGPKEGGLGVNGTAVASAMTSLSDSHMLSLSQ 300
Db 235 VVHGEKEXILYAREMALEFNLPEVLGPEGLGVNGTAVASAMTSLSDSHMLSLSQ 294
QY 301 ALTALTEAMVGOQGSFAPFIHDVCRPHQOVEVARNIRTLSSGSSFAVEHEEVKXKD 360
Db 295 SLTAMTEAMVGHAGSFPHPLHDVTRPHPTQIEVAGNIRKLLEGRFAVHHEEVKXKD 354
QY 361 EGLIKODRYPLEPRTSPQIFGLPVEDMMAVSTLSLE-NNTTNDNPLDVENKQTRAGSNFQ 419
Db 355 EGLIKODRYPLEPRTSPQIFGLPVEDMMAVSTLSLE-NNTTNDNPLDVENKQTRAGSNFQ 414
QY 420 ASAASISMEKTRIALALIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDHI 479
Db 415 AAAAVANTMEKTRILGIAQIGKLNFTQCTELNNAANRGLPSCLAEDPSLNYHGKGLDHI 474
QY 480 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAYVASELGLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534
QY 540 VDLRAAMELDPKKQOPPLPPTLLOOHLG---TGLDV-NALALEVKKALKRLEOQTTTYLE 595
Db 535 IDLRAXEFBPKKQOPPAIVSLIDQFSGAMTGSNLRDELVEKYNKTLAKLEQNTSIDLV 594
QY 596 PRMHDAFSFAGTVEVLST---SLSLAAVANAKVAAESAIISLTQVRETFMSAATS 655
Db 595 PRMHDAFSFAGTVEVLST---SLSLAAVANAKVAAESAIISLTQVRETFMSAATS 651
QY 656 APAHAIVSPRTVLYSPFREELGVQARRGDVGVQOETIGSNVSRITYEAIKQGRINHL 715
Db 652 SPALSYLSPTQILYAFVREELGVKARRGDVFLGQEVYIGSNVSKITYEAIKGRINHL 711
QY 716 VKMLA 720
Db 712 LKMLA 716
RESULT 7
US-09-627-216a-10
; Sequence 10, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
```

```
/ APPLICANT: Sariastani, Sima F
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei
/ APPLICANT: Vannelli, Todd
/ APPLICANT: Gatenby, Anthony
/ TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
/ FILE REFERENCE: BC1009 US NA
/ CURRENT APPLICATION NUMBER: US/09/627,216A
/ PRIORITY FILING DATE: 2000-07-27
/ PRIOR APPLICATION NUMBER: 60/147,719
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: mutant from
US-09-627-216A-10

Query Match      72.4%; Score 2612; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 4e-247;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY      1  MARSLSLATTLANGFTNSHAAPTSAAGPTSAARPTFGLDGHAHQSOLEYVELLSD 60
      1  MARSLSLSHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTYQYVDIVKMLAA 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      61  PTDVVELSGSLTVYADVGAARKGRVRYVNDDEIRAVDSYDFLKAQONSYYGVT 120
      55  PTDSTLELDGSLNIGDVYSAARKGRPVRYVNDSEIRKIDSVFELNSQLSMSYGYTT 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121  GFGGSADRTTEDAVSIQKALIEHQLCGVTPPTSXSFSVGRGIENTLPLEYVRGAMVIRVN 180
      115  GFGGSADRTTEDAISIQKALLEHQLCGVLPSSFDSFRLGRGIENSLPLEYVRGAMTIRVN 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      181  SLTRGSAVRLVYLEALTNFLNRITPIYPLRGSTIASGDSLPSLYIGATIGHPDVYH 240
      175  SLTRGSAVRLVYLEALTNFLNRITPIYPLRGSTIASGDSLPSLYIAAISGHPDSKYH 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      241  VHEGTEKIMFAREASLFGLEAVYLGPEKGLGVNGTAVASAMATLSLHDSHMLSLISQ 300
      235  VVHEGKEKILYAREAMALFNLBPVVLGPEKGLGVNGTAVASAMATLSLHDSHMLSLISQ 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      301  ALTALVTEAMVGOQGSFAPIFDVCRPHGQYEVARNITTLSSGSFAVEHEEVKXKD 360
      295  SLTAMTVEAMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKLEGSRFVHHEEVKXKD 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      361  EGIIRDORYPLRTSPQFLGPIVEDMMHAYSTLSLE--NNTTNDNPLLDYENKOTAHGNGFQ 419
      355  EGIIRDORYPLRTSPQWLGPIVSDLIHAAVLTIEAGSTTNDNPLLDYENKTSHHGNGFQ 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      420  ASAVSISMEXTRLALALIGKLNFTOCTELNLNAANRGIPSCIAADPSLNVHGKGLDHI 479
      415  AAIVANTMEXTRLGLAQIGKLNFTQTEMLNAGNRGIPSCIAADPSLSYHCKGLDIAA 474
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      480  AAASSELGHIANVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTQA 539
      475  AATTSLEGLIANPVTHVQPAEMANOAVNSLALISARPTTSSNDVLSLLATHLYCTQA 534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      540  VDIRAMELDFKQFDPDLLPTLLQOHLG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
      535  IDLRATEFEFKQFGALIVSLIDQHGSGAMTSNLRDELVEKVNNTLAKRLEQNTSYDLV 594
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      596  PRMHDAFSYATGVVELLSSSPSANVTTLAVNNAMKYASAEKALISTREVRKRFMOTPSQ 655
      595  PRMHDAFSYATGVVELLSSSPSANVTTLAVNNAMKYAAAESALISTROYRETFMNASTS 651
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      656  APAHAYISPTRIYAFVVEBELGVCAARRGDVFGVGOQETIGSNVRIEAKIDGINVYL 715
      652  SPHLSYISPTRIYAFVVEBELGVCAARRGDVILGQGEVITIGSNVKITIEAKISGINVYL 711
```

```
QY      716  VKMLA 720
      : : : : :
DB      712  LKMLA 716

RESULT 8
US-09-765-873A-10
/ Sequence 10, Application US/09765873A
/ Patent No. 6521748
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Xiao-Song
/ TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
/ FILE REFERENCE: BC1009 US CIP
/ CURRENT APPLICATION NUMBER: US/09/765,873A
/ PRIORITY FILING DATE: 2001-01-19
/ PRIOR APPLICATION NUMBER: US 09/627,216
/ PRIOR FILING DATE: 2000-07-27
/ PRIOR APPLICATION NUMBER: US 60/147,719
/ PRIOR FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodocorulia glutinis
US-09-765-873A-10

Query Match      72.4%; Score 2612; DB 4; Length 716;
Best Local Similarity 72.0%; Pred. No. 4e-247;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY      1  MARSLSLATTLANGFTNSHAAPTSAAGPTSAARPTFGLDGHAHQSOLEYVELLSD 60
      1  MARSLSLSHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTYQYVDIVKMLAA 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      61  PTDVVELSGSLTVYADVGAARKGRVRYVNDDEIRAVDSYDFLKAQONSYYGVT 120
      55  PTDSTLELDGSLNIGDVYSAARKGRPVRYVNDSEIRKIDSVFELNSQLSMSYGYTT 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      121  GFGGSADRTTEDAVSIQKALIEHQLCGVTPPTSXSFSVGRGIENTLPLEYVRGAMVIRVN 180
      115  GFGGSADRTTEDAISIQKALLEHQLCGVLPSSFDSFRLGRGIENSLPLEYVRGAMTIRVN 174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      181  SLTRGSAVRLVYLEALTNFLNRITPIYPLRGSTIASGDSLPSLYIGATIGHPDVYH 240
      175  SLTRGSAVRLVYLEALTNFLNRITPIYPLRGSTIASGDSLPSLYIAAISGHPDSKYH 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      241  VHEGTEKIMFAREASLFGLEAVYLGPEKGLGVNGTAVASAMATLSLHDSHMLSLISQ 300
      235  VVHEGKEKILYAREAMALFNLBPVVLGPEKGLGVNGTAVASAMATLSLHDSHMLSLISQ 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      301  ALTALVTEAMVGOQGSFAPIFDVCRPHGQYEVARNITTLSSGSFAVEHEEVKXKD 360
      295  SLTAMTVEAMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKLEGSRFVHHEEVKXKD 354
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      361  EGIIRDORYPLRTSPQFLGPIVEDMMHAYSTLSLE--NNTTNDNPLLDYENKOTAHGNGFQ 419
      355  EGIIRDORYPLRTSPQWLGPIVSDLIHAAVLTIEAGSTTNDNPLLDYENKTSHHGNGFQ 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      420  ASAVSISMEXTRLALALIGKLNFTOCTELNLNAANRGIPSCIAADPSLNVHGKGLDHI 479
      415  AAIVANTMEXTRLGLAQIGKLNFTQTEMLNAGNRGIPSCIAADPSLSYHCKGLDIAA 474
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      480  AAASSELGHIANVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTQA 539
      475  AATTSLEGLIANPVTHVQPAEMANOAVNSLALISARPTTSSNDVLSLLATHLYCTQA 534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      540  VDIRAMELDFKQFDPDLLPTLLQOHLG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
      535  IDLRATEFEFKQFGALIVSLIDQHGSGAMTSNLRDELVEKVNNTLAKRLEQNTSYDLV 594
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      596  PRMHDAFSYATGVVELLSSSPSANVTTLAVNNAMKYASAEKALISTREVRKRFMOTPSQ 655
```

[illegible]

RESULT 9

US-09-765-873A-38
Sequence 38, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BC1009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 38
LENGTH: 716
TYPE: PR1
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: UNSURE
LOCATION: (502)..(502)
OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
NAME/KEY: UNSURE
LOCATION: (540)..(540)
OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

Query Match 72.2%; Score 2607; DB 4; Length 716;

Best Local Similarity 71.9%; Pred. No. 1.3e-246,
Matches 521; Conservative 85; Mismatches 105; Indels 14; Gaps 5

QY	1	MASLSLSLATTLLANGFNGTNGSHAAPTKYASAPTSALRRTPGDGHAAHOSQLEIVQELLSD	60	
		1	MASSLSLSISSIFANG-----VASAKQANNGASTYLAAGSHLPITQVOTQVDIYERKLLAA	54
Db				
QY	61	PTDDVVELSGSYLTVRDVVGAAARKGRVAVONDDIIRARVDKSYDFLKAOLONSVCVTT	120	
Db	55	PTSTSTELDGYSLNLDGVYSAARKGRPRVVKDSDIRSKIDSKVEFLRSQLSNVGVTT	114	
QY	121	GGGASDTRFEDVSLQKALIEHOCGTPXTXSFSVGRGLKNTLPLEYVGRGMVTRVN	180	
Db	115	GGGASDTRFEDVSLQKALIEHOCGLVPSFDSFRIGRGLKNSLPLEVVRGMATTRVN	174	
QY	181	SLTRGHSAYELVYLEALTNFLNHRITPIVPLEGSIASAGDLSPLYAGALITGHPDVKN	240	
Db	175	SLTRGHSAYELVYLEALTNFLNHRITPIVPLRGTISAGDLSPLYIAALISGHPDSKH	234	
QY	241	VLEHGEGKIMFAEALSLFGLFAVVLGPREGGLVNGVAVASAMATLSLHDSHMLSLLSQ	300	
Db	235	VVEHGEGKILYAEAAALFNLEBPVVLGPREGGLVNGVAVASAMATLALHDAHMLSLLSQ	294	
QY	301	ALFALVLEAMVGGQGSFAFPIHDVCPREHPGQVEVARNRTLLSSSFAVEHEEVEVKXD	360	
Db	295	SLFAMVLEAMVGGAGSFHPIFLHDVTRPHPTQIVAGCNIRKLLBSSRFVHNHEEVEVKXD	354	
QY	361	EGILRODRYPLRTSPQFLGPIVDEMMHAYSTLSLE-NNTTNDNPLDVENKQTAHGNFQ	419	
Db	355	EGILRODRYPLRTSPQFLGPIVSDIILHAAVYLTIDAGOSTDNPILIDVENTSHHGNFQ	414	

[illegible]

RESULT 10

```

US-09-765-873A-36
/ Sequence 36, Application US/09765873A
/ Patent No.6521748
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Xiao-Song
/ TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
/ FILE REFERENCE: BC1009 US CIP
/ CURRENT APPLICATION NUMBER: US/09/765,873A
/ CURRENT FILING DATE: 2001-01-19
/ PRIOR APPLICATION NUMBER: US 09/627,216
/ PRIOR FILING DATE: 2000-07-27
/ PRIOR APPLICATION NUMBER: US 60/147,719
/ PRIOR FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Microsoft Office 97

```

```

; SEQ ID NO 36
;
; LENGTH: 716

```

```

? TYPE: PR1
? ORGANISM: mutant from Rhodotorula glutinis
? FEATURE:
? NAME/KEY: UNSURE
? LOCATION: (181)..(181)
? OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
? NAME/KEY: UNSURE
? LOCATION: (235)..(235)
? OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
? NAME/KEY: UNSURE
? LOCATION: (540)..(540)
? OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
? US-09-765-873A-36

```

Query Match 72.1%; Score 2603; DB 4; Length 716;

Best Local Similarity 71.7%; Pred. No. 3.1e-246;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5.

Qy 1 MAFSLDLSIATTLANGFTNGSHAPPTKSAAGPTSLARRTPGLDGHAAOSOLEIVOEILLSD 60
1 MAFSLDLSISHPANG-----VASAKQAVNAGNSTLTAAGSHLPTTQTVGVDIYEKTLAA 54
Db
Qy 61 PTTDVVELSGYSITVRDVVGAARKGRVRVONDEIRADVKSVDPLKAOLQNSVYGVT 120
55 PTTSTLELDGYSILNLDVVSAARKGRPVRVSDSEIRSKIRKSYEFLRISOLSMSVYGVT 114
Db
Qy 121 GFGSGADRTEDAVSLQKLLIHHOICGTPPSSXSFSGRGLENTLPLEVVRGMVTRVN 180
115 GFGSGADRTEDAVSLQKLLIHHOICGVLPSSFSGFRGRLENTLPLEVVRGMVTRVN 174

```

Query Match: 72.1%; Score 2603; DB 4; Length 716;
Best Local Similarity 71.7%; Pred. No. 3, 1e-246;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5,

QY      1 MAPSDLSLATTLANGFTNGSHAAPTKSAAGPTSAALRTPTGLDGHAAHSGOLEIVQELLSD 60
DB      1 MAPSDLSLSHSPFANG-----VASAQAQVAGKASTNLAIVAGSHLPTTQVQVDDIVKMLAA 54
QY      61 PDDVDVEISGLSVLRVDVGAAKGRVRVQNDDELRADVDSVPLKAOQNSYVGT 120
DB      55 PDDSTLEIDGSLWGDVYSAARKGRVAKVQKDELRKIDSVSEFLSOLSMSYVGT 114
QY      121 GFGSGADRTEDAVSLQKALIEHQLCGVPTPSXSFVSFVGRLENTLPLEVYRGAMVIRN 180
DB      115 GFGSGADRTEDALSLQKALIEHQLCGVLPSPDFXRLGRGLENSLPLEVYRGAMTIRVN 174
QY      181 SLTRGSHSVRLVLEALTMFLNRRITPYVLRGSIASGDSLPLSYAGATGHPDVYH 240
DB      175 SLTRGSHSVRLVLEALTMFLNRRITPYVLRGTISASGDSLPLSYIAALSGHDSKVH 234
QY      241 VLHSTGKIMFARPAISLFGLEAVLVGPKSGGLVNGTIVASASMATLSLHDSHMLSLSQ 300
DB      235 VVHEKEKILVARBAMMLFMLEPVLVGPKSGGLVNGTIVASASMATLALHDAHMLSLSQ 294
QY      301 ALTALTEAMTGGCGGSAPTFIHDVCRPHQGVFAANITLLSGSFVAVEHEEVKVD 360
DB      295 SLTAMTAVMAGHAGSPHPLHDVTRPHPTQIEVAGNIRKLEGRFVHHEEVKVD 354
QY      361 EGLIKODRYPRTSPQPLGPIVDDMMHAYSTLSL-NNTTNDPLDYEYKOTAHGKVFQ 419
DB      355 EGLIKODRYPRTSPQPLGPIVDDLHAHAVLTIEAGSGTTNPLIDYENKSHHGKVFQ 414
QY      420 ASAVSISMETRLALALIGKLNFTQCTELINAAANGDPSCLAEDPSLNYHGKLDIHI 479
DB      415 AAAYAMTMEKTRLGLAQIGKLNFTQCTELNMGNFGDPSCLAEDPSLSYHGKLDIAA 474
QY      480 AAAYSEGLHANPTTTPVQPAEMGNQAVNSLAISSARPAENDVLSLASHVCTLOA 539
DB      475 AAAYSEGLHANPTTTPVQPAEMGNQAVNSLAISSARPTESNDVLSLASHVCTLOA 534
QY      540 VDLRAMELDFFKQCPDLLPTLLOOHLG---TGLDV-NALALEVKALNKRLEQTTYYLE 595
DB      535 IDLRXPEPFKKQCGPAIVSLIDQHGSGAMTGSNLDELVEKVNITLAKRLEQTNSTYLV 594
QY      596 PRMDHAFSALGTYVELLSSPSBANVTLRVANAMKAKSAEKALSLTRERVRNFQTPSSQ 655
DB      595 PRMDHAFSALGTYVELLSSPSBANVTLRVANAMKAKSAEKALSLTRQVRETFVSAASTS 651
QY      656 AAPAHYLSFPRRVLYSVFVEBELGVQARGGVFGVQOEITGSNVRIVEALIKDGINVL 715
DB      653 SPALSYLSPRQLIYXAVREBELGVKARGGVFLGKQEVITGSNVSKITFALKSGINNVL 711
QY      716 VKMLA 720
DB      712 LKMLA 716

RESULT 12
US-09-765-873A-34
; Sequence 34, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: B-OPRODUCTION OF PARA-HYDROXYINNAMIC ACID
; FILE REFERENCE: BCI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97

```



```
/ SEQ ID NO 34
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodotorula glutinis
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (126)..(126)
/ OTHER INFORMATION: X= Gly, Ala, Ser, Thr
/ NAME/KEY: UNSURE
/ LOCATION: (138)..(138)
/ OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
/ NAME/KEY: UNSURE
/ LOCATION: (540)..(540)
/ OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-34
```

Query Match 72.0%; Score 2599; DB 4; Length 716;

Best Local Similarity 71.7%; Pred. No. 7,6e-246; Mismatches 106; Indels 14; Gaps 5;

Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

```
QY 1 MAPSLDLSIATTLANGFTNGSHAAPTSAAGPTSAURTPGDLGHAHQSOLEIYQELISD 60
DB 1 MAPSLDLSISHFANG-----VASAKQAVNGASTNLAVAGSHLPTQTQYQYDIYVKMLAA 54
QY 61 PTDDVVELSGYSLTYRDVVGAAKGRVRYQNDDEIRAVDKSVDFLKAQIQNSVYGVTT 120
DB 55 PTDTSTLELDGYSINLDVVSAAKGRPVRYKQSDERISKIDKSVFELKSQLSMSVYGVTT 114
QY 121 GFGSADRTEDAVSLQKALIEHQLCGVTPTSXSFSGVSGLENTLPLEVVRGAMVIRVN 180
DB 115 GFGSADRTREXALISIQKALIEHQLCGVTPSSFDRLRGLENTLPLEVVRGAMVIRVN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPVLPGSISASGDLSPSYIAGATGHPDVKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPVLPGSISASGDLSPSYIAGATGHPDVKH 234
QY 241 VLHEGTEKIMFAREALISFGLAEVVLGPKRGLGVNGTAVASAMATLSLHSHMLSLISQ 300
DB 235 VLHEGTEKILYAREAMALFNLPEVVLGPKRGLGVNGTAVASAMATLSLHSHMLSLISQ 294
QY 301 ALATLVEAAVVGQGSFAPFIHDVCRPHRQVEVANRITRLSGSSFAVEHEEVKXKD 360
DB 295 SLTAMTVEAAVVGAGSFHPLHDVTPRHPTQIEVAGNIRKLEGSFVHHEEVKXKD 354
QY 361 EGIIRDORVPLRTSPQPLPIVEDMMAHASTLSLE--NNTTNDNPLDLENQIAGAGFQ 419
DB 355 EGIIRDORVPLRTSPQPLPIVEDMMAHASTLSLE--NNTTNDNPLDLENQIAGAGFQ 414
QY 420 ASAVSISMEKIRLALALIGLNFTQCTELNAAANRGLPSCIAADPSILNHYGQLDIHI 479
DB 415 AAVALNTMEKTRIGLAIQIGLNFTQCTELNAAANRGLPSCIAADPSILNHYGQLDIHI 474
QY 480 AAYASISGLHANPYTTFVOPAEKGNQAVNSLALISARTAEANDVLSLLASHLYCTLOA 539
DB 475 AAYTSEIGHLANPYTHVOPAEKGNQAVNSLALISARTAEANDVLSLLASHLYCTLOA 534
QY 540 VDIRAMELDEKQFDPLPTLLQOHLG--TGIDY--NALALEVKKALKRLEQTTVDLE 595
DB 535 IDIRAKEFEFKQFGAIVSLIDQHFSGAMTGSNLRDELIVERNKTLAKRLEQNSIDLV 594
QY 596 PRWHDASFVATGTVELLSSSPBANVTITAVANAKVASHAKAISITREVRNRFWQTPSSQ 655
DB 595 PRWHDASFVATGTVELLSSSPBANVTITAVANAKVASHAKAISITREVRNRFWQTPSSQ 651
QY 656 APAAATLSPRTVLVFFVEEELGVQARBGDVFGVQOETIGSNVSRIVEALIKDRIHNVL 715
DB 652 SPLASTUSPRTQTLVFFVEEELGVQARBGDVFGVQOETIGSNVSRIVEALIKDRIHNVL 711
QY 716 VKMLA 720
DB 712 VKMLA 716
```

RESULT 13

US-09-765-873A-33

Sequence 33, Application US/05765873A

Patent No. 6521748

GENERAL INFORMATION:

APPLICANT: tang, Xiao-Song

TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID

FILE REFERENCE: BC1009 US CIP

CURRENT APPLICATION NUMBER: US/09/765, 873A

CURRENT FILING DATE: 2001-01-19

PRIOR APPLICATION NUMBER: US 09/627, 216

PRIOR FILING DATE: 2000-07-27

PRIOR APPLICATION NUMBER: US 60/147, 719

PRIOR FILING DATE: 1999-08-06

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Microsoft Office 97

SEQ ID NO 33

LENGTH: 716

TYPE: PRT

ORGANISM: mutant from Rhodotorula glutinis

FEATURE:

NAME/KEY: UNSURE

LOCATION: (126)..(126)

OTHER INFORMATION: X= Gly, Ser, Ala, Thr

NAME/KEY: UNSURE

LOCATION: (138)..(138)

OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys

NAME/KEY: UNSURE

LOCATION: (149)..(149)

OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly

NAME/KEY: UNSURE

LOCATION: (181)..(181)

OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly

NAME/KEY: UNSURE

LOCATION: (198)..(198)

OTHER INFORMATION: X= Asp, Asn, Glu, Gln

NAME/KEY: UNSURE

LOCATION: (202)..(202)

OTHER INFORMATION: X= Val, Met, Leu, Cys

NAME/KEY: UNSURE

LOCATION: (235)..(235)

OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro

NAME/KEY: UNSURE

LOCATION: (502)..(502)

OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro

NAME/KEY: UNSURE

LOCATION: (540)..(540)

OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly

US-09-765-873A-33

Query Match

71.4%; Score 2579; DB 4; Length 716;

Best Local Similarity 71.0%; Pred. No. 7e-244; Mismatches 110; Indels 14; Gaps 5;

Matches 515; Conservative 86; Mismatches 110; Indels 14; Gaps 5;

```
QY 1 MAPSLDLSIATTLANGFTNGSHAAPTSAAGPTSAURTPGDLGHAHQSOLEIYQELISD 60
DB 1 MAPSLDLSISHFANG-----VASAKQAVNGASTNLAVAGSHLPTQTQYQYDIYVKMLAA 54
QY 61 PTDDVVELSGYSLTYRDVVGAAKGRVRYQNDDEIRAVDKSVDFLKAQIQNSVYGVTT 120
DB 55 PTDTSTLELDGYSINLDVVSAAKGRPVRYKQSDERISKIDKSVFELKSQLSMSVYGVTT 114
QY 121 GFGSADRTEDAVSLQKALIEHQLCGVTPTSXSFSGVSGLENTLPLEVVRGAMVIRVN 180
DB 115 GFGSADRTREXALISIQKALIEHQLCGVTPSSFDRLRGLENTLPLEVVRGAMVIRVN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPVLPGSISASGDLSPSYIAGATGHPDVKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPVLPGSISASGDLSPSYIAGATGHPDVKH 234
QY 241 VLHEGTEKIMFAREALISFGLAEVVLGPKRGLGVNGTAVASAMATLSLHSHMLSLISQ 300
DB 235 VLHEGTEKILYAREAMALFNLPEVVLGPKRGLGVNGTAVASAMATLSLHSHMLSLISQ 294
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:07:54 ; Search time 92.6141 Seconds
(without alignments)
2493.103 Million cell updates/sec

Title: US-09-939-408a-13

Perfect score: 3610
1 MAPSLDLSLATTLANGFTNGS.....RIYKAIKDGRIHVLVKKLA 720

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3608	99.9	720	9 US-09-939-408a-13	Sequence 13, Appl
2	2770	76.7	726	9 US-09-939-408a-21	Sequence 21, Appl
3	2730	75.6	720	9 US-09-939-408a-30	Sequence 30, Appl
4	2614	72.4	716	9 US-09-765-873a-8	Sequence 8, Appl
5	2614	72.4	716	9 US-09-939-408a-19	Sequence 19, Appl
6	2614	72.4	716	14 US-10-138-970a-4	Sequence 4, Appl
7	2614	72.4	716	14 US-10-188-523b-8	Sequence 8, Appl
8	2614	72.4	716	15 US-10-374-366-2	Sequence 2, Appl
9	2614	72.4	716	15 US-10-439-478-2	Sequence 35, Appl
10	2613	72.4	716	9 US-09-765-873a-35	Sequence 19, Appl
11	2613	72.4	716	14 US-10-138-970a-19	Sequence 35, Appl
12	2613	72.4	716	14 US-10-188-523b-35	Sequence 35, Appl
13	2613	72.4	716	15 US-10-439-478-15	Sequence 15, Appl
14	2612	72.4	716	9 US-09-765-873a-10	Sequence 10, Appl
15	2612	72.4	716	14 US-10-138-970a-24	Sequence 24, Appl

16	2612	72.4	716	14 US-10-188-523b-10	Sequence 10, Appl
17	2612	72.4	716	15 US-10-374-366-4	Sequence 4, Appl
18	2609	72.3	716	15 US-10-374-366-24	Sequence 24, Appl
19	2608	72.2	716	15 US-10-374-366-22	Sequence 22, Appl
20	2607	72.2	716	9 US-09-765-873a-38	Sequence 38, Appl
21	2607	72.2	716	14 US-10-138-970a-22	Sequence 22, Appl
22	2607	72.2	716	14 US-10-188-523b-38	Sequence 38, Appl
23	2607	72.2	716	15 US-10-374-366-16	Sequence 16, Appl
24	2607	72.2	716	15 US-10-439-478-18	Sequence 18, Appl
25	2606	72.2	716	15 US-10-374-366-8	Sequence 8, Appl
26	2605	72.2	716	15 US-10-374-366-10	Sequence 10, Appl
27	2604	72.1	716	15 US-10-374-366-14	Sequence 14, Appl
28	2603	72.1	716	9 US-09-765-873a-36	Sequence 36, Appl
29	2603	72.1	716	9 US-09-765-873a-37	Sequence 37, Appl
30	2603	72.1	716	14 US-10-138-970a-20	Sequence 20, Appl
31	2603	72.1	716	14 US-10-138-970a-21	Sequence 21, Appl
32	2603	72.1	716	14 US-10-188-523b-36	Sequence 36, Appl
33	2603	72.1	716	14 US-10-188-523b-37	Sequence 37, Appl
34	2603	72.1	716	15 US-10-439-478-16	Sequence 16, Appl
35	2603	72.1	716	15 US-10-439-478-17	Sequence 17, Appl
36	2601	72.0	716	15 US-10-374-366-12	Sequence 12, Appl
37	2599	72.0	716	9 US-09-765-873a-34	Sequence 34, Appl
38	2599	72.0	716	14 US-10-138-970a-18	Sequence 18, Appl
39	2599	72.0	716	14 US-10-188-523b-34	Sequence 34, Appl
40	2599	72.0	716	15 US-10-374-366-18	Sequence 18, Appl
41	2599	72.0	716	15 US-10-439-478-14	Sequence 14, Appl
42	2598	72.0	716	15 US-10-374-366-6	Sequence 6, Appl
43	2598	72.0	716	15 US-10-374-366-20	Sequence 20, Appl
44	2579	71.4	716	9 US-09-765-873a-33	Sequence 33, Appl
45	2579	71.4	716	14 US-10-138-970a-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-939-408a-13
Sequence 13, Application US/0939408a
Patient No. US20020102712H1
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
APPLICANT: Koestera, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
TITLE OF INVENTION: Using Same
FILE REFERENCE: 294/9/500NSCA
CURRENT APPLICATION NUMBER: US/09/939,408A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodocorula graminis
FEATURES:
NAME/KEY: unsure
LOCATION: (153)
OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408a-13

Query Match 99.9%, Score 3608, DB 9, Length 720;
Best Local Similarity 100.0%, Pred. No. 0;
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSLDLSLATTLANGFTNGSHAAPTSAAGPTSAURRTPTGIDGHAHOSQLEIVGELISD 60
DB 1 MAPSLDLSLATTLANGFTNGSHAAPTSAAGPTSAURRTPTGIDGHAHOSQLEIVGELISD 60
QY 61 PTDVVVLSGSLTVADVVGAARKGRVAVQVNDDEIRARVDKSVDFLKAQJLONSIVGVTT 120

```

Db      61 PTDVVLSSGLSVLRVVGAAKGRVVRQNDDEIRAVDVSVPDLKQLONSVYGVTT 120
Qy      121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSGRGENTLPLEVVGAMVIRN 180
Db      121 GFGGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSGRGENTLPLEVVGAMVIRN 180
Qy      181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDLSPSYIAGATGHPDVKYH 240
Db      181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDLSPSYIAGATGHPDVKYH 240
Qy      241 VLNHGTEKIMFARBAISLFGLEAVVIGPKEGGLVNGTAVSASMATLSLHDSHMLSLSSQ 300
Db      241 VLNHGTEKIMFARBAISLFGLEAVVIGPKEGGLVNGTAVSASMATLSLHDSHMLSLSSQ 300
Qy      301 ALTALTYEAMVGQGSFAPFIHDVCRPHQGVAVANIRTLISGSSFAVEHEEVKXVD 360
Db      301 ALTALTYEAMVGQGSFAPFIHDVCRPHQGVAVANIRTLISGSSFAVEHEEVKXVD 360
Qy      361 EGIIRORRYPLRTSPQPLGVLVEDMMAHASTLSLENNITTDNPILDVEMKQTAHGNFQA 420
Db      361 EGIIRORRYPLRTSPQPLGVLVEDMMAHASTLSLENNITTDNPILDVEMKQTAHGNFQA 420
Qy      421 SAVSISMEKTRIALALIGKLNFTQCTELNPAANRGLPSCLAEDPSLNYHGKGLDIIIA 480
Db      421 SAVSISMEKTRIALALIGKLNFTQCTELNPAANRGLPSCLAEDPSLNYHGKGLDIIIA 480
Qy      481 AVASELGHANPVTTFVQPAEMGNQAVNSIALISARRTAEADVLSLLASHLYCTLOAV 540
Db      481 AVASELGHANPVTTFVQPAEMGNQAVNSIALISARRTAEADVLSLLASHLYCTLOAV 540
Qy      541 DLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNALAEVKKALNKELEQTTTDLDEPRHND 600
Db      541 DLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNALAEVKKALNKELEQTTTDLDEPRHND 600
Qy      601 AFSYATGTVEELLSSSPSANVTLTAVANMKVSAEKAISLTREVNRFWQTPSSQAPAAH 660
Db      601 AFSYATGTVEELLSSSPSANVTLTAVANMKVSAEKAISLTREVNRFWQTPSSQAPAAH 660
Qy      661 YLSRTTRVXSVFVEELGVQARRGDVFGVQOETIGSNVSRITYEALIKOGRINHYLVKMA 720
Db      661 YLSRTTRVXSVFVEELGVQARRGDVFGVQOETIGSNVSRITYEALIKOGRINHYLVKMA 720

```

RESULT 2

```

US-09-939-408a-21
; Sequence 21, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koostera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)..(719)
; OTHER INFORMATION: "Xaa"means any amino acid; "Xaa"means no consensus at that
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19

```

US-09-939-408a-21

Query Match 76.7%; Score 2770; DB 9; Length 726;
 Best Local Similarity 77.7%; Pred. No. 5,8e-243;
 Matches 564; Conservative 40; Mismatches 116; Indels 6; Gaps 3;

```

Qy      1 MAPSLDLSLTTLANGFTNGSHAPFKSAAGPTSLARPGLDGHAAHQSLEIYQELISLD 60
Db      1 MAPSLDLSLTTLANGFTNGSHAPFKSAAGPTSLARPGLDGHAAHQSLEIYQELISLD 60
Qy      61 P-TDVVELSGSLTFVRDVVGAARKGRVVRQNDDEIRAVDVSVPDLKQLONSVYGVTT 119
Db      61 P-TDVVELSGSLTFVRDVVGAARKGRVVRQNDDEIRAVDVSVPDLKQLONSVYGVTT 119
Qy      120 TFGGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSGRGENTLPLEVVGAMVIRN 179
Db      120 TFGGSADTRTEDAVSLQKALIEHQLCVPTXSXSSVSGRGENTLPLEVVGAMVIRN 179
Qy      180 NSLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDLSPSYIAGATGHPDVKY 239
Db      180 NSLTRGSAVRLVLEALTNFLNHRITPIVPLRGSSISASGDLSPSYIAGATGHPDVKY 239
Qy      240 HVLHGTEKIMFARBAISLFGLEAVVIGPKEGGLVNGTAVSASMATLSLHDSHMLSLSS 299
Db      240 HVLHGTEKIMFARBAISLFGLEAVVIGPKEGGLVNGTAVSASMATLSLHDSHMLSLSS 299
Qy      300 QALITATYAMVGQGSFAPFIHDVCRPHQGVAVANIRTLISGSSFAVEHEEVKXVD 359
Db      300 QALITATYAMVGQGSFAPFIHDVCRPHQGVAVANIRTLISGSSFAVEHEEVKXVD 359
Qy      360 QALITATYAMVGQGSFAPFIHDVCRPHQGVAVANIRTLISGSSFAVEHEEVKXVD 418
Db      360 QALITATYAMVGQGSFAPFIHDVCRPHQGVAVANIRTLISGSSFAVEHEEVKXVD 418
Qy      419 QASAVSISMEKTRIALALIGKLNFTQCTELNPAANRGLPSCLAEDPSLNYHGKGLDII 478
Db      419 QASAVSISMEKTRIALALIGKLNFTQCTELNPAANRGLPSCLAEDPSLNYHGKGLDII 478
Qy      479 IAAVASELGHANPVTTFVQPAEMGNQAVNSIALISARRTAEADVLSLLASHLYCTLO 538
Db      479 IAAVASELGHANPVTTFVQPAEMGNQAVNSIALISARRTAEADVLSLLASHLYCTLO 538
Qy      539 AVDLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNA---LAEVKKALNKELEQTTTYDL 594
Db      539 AVDLRAEMELDFKKQDPPLPTLLOHIGTGGLDVNA---LAEVKKALNKELEQTTTYDL 594
Qy      595 EPRMWDASVATGTVEELLSSSPSANVTLTAVANMKVSAEKAISLTREVNRFWQTPSS 654
Db      595 EPRMWDASVATGTVEELLSSSPSANVTLTAVANMKVSAEKAISLTREVNRFWQTPSS 654
Qy      601 EPRMWDASVATGTVEELLSSSPSANVTLTAVANMKVSAEKAISLTREVNRFWQTPSS 660
Db      601 EPRMWDASVATGTVEELLSSSPSANVTLTAVANMKVSAEKAISLTREVNRFWQTPSS 660
Qy      655 QAPAAVYLSPTTRVXSVFVEELGVQARRGDVFGVQOETIGSNVSRITYEALIKOGRINHY 714
Db      655 QAPAAVYLSPTTRVXSVFVEELGVQARRGDVFGVQOETIGSNVSRITYEALIKOGRINHY 714
Qy      715 LVKMA 720
Db      721 LVKMA 726

```

RESULT 3

```

US-09-939-408a-30
; Sequence 30, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koostera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693

```

```

PRIORITY FILING DATE: 2000-07-24
PRIORITY APPLICATION NUMBER: PCT/US01/23270
PRIORITY FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 30
LENGTH: 720
TYPE: PR
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Preferred
OTHER INFORMATION: Theoretical sequence based in part on SEQ ID NO:20
NAME/KEY: misc_feature
LOCATION: (5)..(5)
OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
NAME/KEY: misc_feature
LOCATION: (12)..(12)
OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
NAME/KEY: misc_feature
LOCATION: (16)..(16)
OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
NAME/KEY: misc_feature
LOCATION: (17)..(17)
OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
NAME/KEY: misc_feature
LOCATION: (19)..(19)
OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
NAME/KEY: misc_feature
LOCATION: (20)..(20)
OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
NAME/KEY: misc_feature
LOCATION: (25)..(25)
OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
NAME/KEY: misc_feature
LOCATION: (27)..(27)
OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
NAME/KEY: misc_feature
LOCATION: (28)..(28)
OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
NAME/KEY: misc_feature
LOCATION: (34)..(34)
OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
NAME/KEY: misc_feature
LOCATION: (36)..(36)
OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
NAME/KEY: misc_feature
LOCATION: (39)..(39)
OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
NAME/KEY: misc_feature
LOCATION: (40)..(40)
OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
NAME/KEY: misc_feature
LOCATION: (48)..(48)
OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala,
NAME/KEY: misc_feature
LOCATION: (54)..(54)
OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
NAME/KEY: misc_feature
LOCATION: (56)..(56)
OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
NAME/KEY: misc_feature
LOCATION: (65)..(65)
OTHER INFORMATION: The 'Xaa' at location 65 stands for Glu, Asp, or Val.
NAME/KEY: misc_feature
LOCATION: (66)..(66)
OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
NAME/KEY: misc_feature
LOCATION: (76)..(76)
OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
NAME/KEY: misc_feature
LOCATION: (87)..(87)
OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
NAME/KEY: misc_feature
LOCATION: (93)..(93)
OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
NAME/KEY: misc_feature
LOCATION: (102)..(102)
OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
NAME/KEY: misc_feature
LOCATION: (103)..(103)
OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
NAME/KEY: misc_feature
LOCATION: (109)..(109)
OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
NAME/KEY: misc_feature
LOCATION: (112)..(112)
OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
NAME/KEY: misc_feature
LOCATION: (114)..(114)
OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
NAME/KEY: misc_feature
LOCATION: (117)..(117)
OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
NAME/KEY: misc_feature
LOCATION: (148)..(148)
OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
NAME/KEY: misc_feature
LOCATION: (150)..(150)
OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
NAME/KEY: misc_feature
LOCATION: (153)..(153)
OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
NAME/KEY: misc_feature
LOCATION: (154)..(154)
OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
NAME/KEY: misc_feature
LOCATION: (157)..(157)
OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
NAME/KEY: misc_feature
LOCATION: (183)..(183)
OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
NAME/KEY: misc_feature
LOCATION: (223)..(223)
OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
NAME/KEY: misc_feature
LOCATION: (225)..(225)
OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
NAME/KEY: misc_feature
LOCATION: (237)..(237)
OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala,
NAME/KEY: misc_feature
LOCATION: (239)..(239)
OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
NAME/KEY: misc_feature
LOCATION: (241)..(241)
OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
NAME/KEY: misc_feature
LOCATION: (242)..(242)
OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
NAME/KEY: misc_feature
LOCATION: (246)..(246)
OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
NAME/KEY: misc_feature
LOCATION: (251)..(251)
OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
NAME/KEY: misc_feature
LOCATION: (259)..(259)
OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
NAME/KEY: misc_feature
LOCATION: (305)..(305)
OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
NAME/KEY: misc_feature
```

LOCATION: (319) .. (319)
OTHER INFORMATION: The 'Xaa' at location 319 stands for Pro.
NAME/KEY: misc feature
LOCATION: (346) .. (346)
OTHER INFORMATION: The 'Xaa' at location 346 stands for Lys, Arg, Thr, Gln, Pro, a
OTHER INFORMATION: stop codon, Trp, or Ser.
NAME/KEY: misc feature
LOCATION: (411) .. (411)
OTHER INFORMATION: The 'Xaa' at location 411 stands for Lys, Thr, or Met.
NAME/KEY: misc feature
LOCATION: (421) .. (421)
OTHER INFORMATION: The 'Xaa' at location 421 stands for Thr, Ala, or Ser.
NAME/KEY: misc feature
LOCATION: (457) .. (457)
OTHER INFORMATION: The 'Xaa' at location 457 stands for Leu.
NAME/KEY: misc feature
LOCATION: (458) .. (458)
OTHER INFORMATION: The 'Xaa' at location 458 stands for Pro.
NAME/KEY: misc feature
LOCATION: (466) .. (466)
OTHER INFORMATION: The 'Xaa' at location 466 stands for Pro.
NAME/KEY: misc feature
LOCATION: (487) .. (487)
OTHER INFORMATION: The 'Xaa' at location 487 stands for Gly.
NAME/KEY: misc feature
LOCATION: (493) .. (493)
OTHER INFORMATION: The 'Xaa' at location 493 stands for Val.
NAME/KEY: misc feature
LOCATION: (518) .. (518)
OTHER INFORMATION: The 'Xaa' at location 518 stands for Thr.
NAME/KEY: misc feature
LOCATION: (536) .. (536)
OTHER INFORMATION: The 'Xaa' at location 536 stands for Val.
NAME/KEY: misc feature
LOCATION: (556) .. (556)
OTHER INFORMATION: The 'Xaa' at location 556 stands for Met, Ile, Val, or Leu.
NAME/KEY: misc feature
LOCATION: (557) .. (557)
OTHER INFORMATION: The 'Xaa' at location 557 stands for Ile, Val, or Leu.
NAME/KEY: misc feature
LOCATION: (558) .. (558)
OTHER INFORMATION: The 'Xaa' at location 558 stands for Thr, Ala, or Pro.

Query Match 75.6%; Score 2730; DB 9; Length 720;

Best Local Similarity 76.5%; Pred. No. 2.5e-239;

Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

QY 1 MASPISLSTATTLANGFTNGSHAAPTKSAAGPTSAARTRPGLDGHAHQSOLIEYOELLSD 60
DB 1 MAASXOSIAITSSXNGKXNXXHAAFXKXKTGATSLXRPXKXLAPEPTQYTDIDIXLAD 60
QY 61 PTDDVVELSGYSLTVADVGAARKGRVRYQNDDEIRARVDSYDFLKAOLQNSYGVTT 120
DB 61 PTDDXXELDGYTLTLKDVGAARKGRVRYQNDDEIRAKIDXXVEFLRXQXNNVYVTT 120
QY 121 GFGGSAADRTEDAVSLQKALIHQOLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRN 180
DB 121 GFGGSAADRTEDAVSLQKALIHQOLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRN 180
QY 181 SLTRGSAVRLVLEALTNFNLHRTPIVPLRGSISASGDISPLSYIGAITGAPDVKH 240
DB 181 SLTRGSAVRLVLEALTNFNLHRTPIVPLRGSISASGDISPLSYIGAITGAPDVKH 240
QY 241 VLHEGTEKINFAREALISLFGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
DB 241 VLHEGTEKINFAREALISLFGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
QY 301 ALTALTVEAMVGGOGSFAPFTHDVCRRPHGQVEVARNIRTLSSGSFVHEEVEVKYKD 360
DB 301 ALTALTVEAMVGGOGSFAPFTHDVCRRPHGQVEVARNIRTLSSGSFVHEEVEVKYKD 360
QY 361 EGLRDRYPLRTSPQFLPVEDWMAVSTLSLNNTTNPLLDVENKOTAHGNFOA 420

DB 361 EGLRDRYPLRTSPQFLPVEDWMAVSTLSLNNTTNPLLDVENKOTAHGNFOA 420
QY 421 SAVSISMEXTRLALALIGKLNFTOCTELNNAANRGIPSCLAADPSLNHGGLDHLA 480
DB 421 XAVANTMKRTRLALALIGKLNFTOCTELNNAANRGIPSCLAADPSLNHGGLDHLA 480
QY 481 AVASELGLANVTTFVOPAEQNAVNSIALISARRTAEANDVLSLASHLYCTLOAV 540
DB 481 AVASELGLANVTTFVOPAEQNAVNSIALISARRTAEANDVLSLASHLYCTLOAV 540
QY 541 DIRAMELDFKQFDBLLPTLLQOHGGLDVNALALEVKKALNRLIEQTTTYLEPRWD 600
DB 541 DIRAMELDFKQFDBLLPTLLQOHGGLDVNALALEVKKALNRLIEQTTTYLEPRWD 600
QY 601 AFSVATGTVVELLSSSPANVTLVANAMKVASAEKASITREVRNRFMQTPSSQAPHA 660
DB 601 AFSVATGTVVELLSSSPANVTLVANAMKVASAEKASITREVRNRFMQTPSSQAPHA 660
QY 661 YLSPRTYLYGFVEBELGVQARQDVEVQOETIGSNVRIYEAIRDRGRINHYVWMLA 720
DB 661 YLSPRTYLYGFVEBELGVQARQDVEVQOETIGSNVRIYEAIRDRGRINHYVWMLA 720

RESULT 4

US-09-765-873A-8

Sequence 8, Application US/09765873A

Patent No. US20010053847A1

GENERAL INFORMATION:

APPLICANT: Jang, Xiao-Song

TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID

FILE REFERENCE: BC1009 US CIP

CURRENT APPLICATION NUMBER: US/09/765, 873A

CURRENT FILING DATE: 2001-01-19

Prior Application Number: US 09/627, 216

Prior Filing Date: 2000-07-27

Prior Application Number: US 60/147, 719

Prior Filing Date: 1999-08-06

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Microsoft Office 97

SEQ ID NO 8

LENGTH: 716

TYPE: PRK

ORGANISM: Rhodotorula glutinis

US-09-765-873A-8

Query Match 72.4%; Score 2614; DB 9; Length 716;

Best Local Similarity 72.0%; Pred. No. 9.5e-229;

Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY 1 MASPISLSTATTLANGFTNGSHAAPTKSAAGPTSAARTRPGLDGHAHQSOLIEYOELLSD 60
DB 1 MAASXOSIAITSSXNGKXNXXHAAFXKXKTGATSLXRPXKXLAPEPTQYTDIDIXLAD 60
QY 61 PTDDVVELSGYSLTVADVGAARKGRVRYQNDDEIRARVDSYDFLKAOLQNSYGVTT 120
DB 61 PTDDXXELDGYTLTLKDVGAARKGRVRYQNDDEIRAKIDXXVEFLRXQXNNVYVTT 120
QY 121 GFGGSAADRTEDAVSLQKALIHQOLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRN 180
DB 121 GFGGSAADRTEDAVSLQKALIHQOLCGVTPISXSSFSVGRGLENTLPLEVVRGAMTVRN 180
QY 181 SLTRGSAVRLVLEALTNFNLHRTPIVPLRGSISASGDISPLSYIGAITGAPDVKH 240
DB 181 SLTRGSAVRLVLEALTNFNLHRTPIVPLRGSISASGDISPLSYIGAITGAPDVKH 240
QY 241 VLHEGTEKINFAREALISLFGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
DB 241 VLHEGTEKINFAREALISLFGLEAVVLGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
QY 301 ALTALTVEAMVGGOGSFAPFTHDVCRRPHGQVEVARNIRTLSSGSFVHEEVEVKYKD 360
DB 301 ALTALTVEAMVGGOGSFAPFTHDVCRRPHGQVEVARNIRTLSSGSFVHEEVEVKYKD 360
QY 361 EGLRDRYPLRTSPQFLPVEDWMAVSTLSLNNTTNPLLDVENKOTAHGNFOA 420


```

QY      361 EGILRODRYPLRTSPQFLGPIVEDMMAVSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 419
      355 EGILRODRYPLRTSPQFLGPIVEDMMAVSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 414
      420 ASAVSISMKEKRLALAIIGKINFTQCTELLNANRGIPSCLAEDPSLHYGGLDHI 479
      415 AAIVANWMEKRLGIAQIGKINFTQCTELLNANRGIPSCLAEDPSLHYGGLDHI 474
      480 AAIVASIELGLANPVTTFFVQPAEMGNOAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
      475 AATSTELGLANPVTTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCYLOA 534
      540 VDLRAMELDFFKQFDPPLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
      535 IDLRAIEFEFFKQFGPAIVSLIDQHFSGAMTGSNLRDELVEKNKTLAKRLEQTNISYDLV 594
      596 PRMHDAFSYATGTVEYELSSPSANVTITAVANAMVSAEKAISLTREVRNRFQOTSSQ 655
      595 PRMHDAFSYATGTVEYELSSPSANVTITAVANAMVSAEKAISLTREVRNRFQOTSSQ 651
      656 APAAVYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHV 715
      652 SPALSTYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHV 711
      716 VKMLA 720
      712 LKMLA 716

RESULT 5
US-09-939-408a-19
; Sequence 19, Application US/09939408a
; Patent No. US20020102712a1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koester, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/50NSCA
; CURRENT APPLICATION NUMBER: US/09/939,408a
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula toruloides
US-09-939-408a-19

Query Match      72.4%; Score 2614; DB 9; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-229;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1 MAPSLDSLATTLANGFTNGSHAAPTSAAGPTSAIRRTPTGLDGHAAHOSQLEIYQELISD 60
      1 MAPSLDSISHSFANG-----VASAKQAVNGASTWLAAGSHLPTTQVTOYDIVERKMLAA 54
      61 PTDDVVELSGYSLTVADVGAARKGRVRVONDEIRARVDKSVDFLKAQONSYGYGTT 120
      55 PTSTLELDGYSLNIGDVVSAARKGRPVAVKSDIRSKIDKSVDFLKAQONSYGYGTT 114
      121 GFGGSADRTTEDAVSLQKALIEHOLCGVTPPSXSFSGRGLENTLPLEVVRGAMVIRVN 180
      115 GFGGSADRTTEDAVSLQKALIEHOLCGVTPPSXSFSGRGLENTLPLEVVRGAMVIRVN 174
      181 SLTRGSHAVLVVLELTNFTNIRIPVPLRGSTISASGDSLPLSYIGATIGHPDVVH 240
      175 SLTRGSHAVLVVLELTNFTNIRIPVPLRGSTISASGDSLPLSYIGATIGHPDVVH 234

```

```

QY      241 VLHEGTEKIMFARREAIISLFGLEAVVLGPKEGGLGVNGTAVASAMATLISLDSHMLSLSQ 300
      235 VVHEGKEKILYAREMAALFNLEPVLPGRKGLGVNGTAVASAMATLISLDSHMLSLSQ 294
      301 AATLTAVAMWQOSSFAPRIHDVCRPHPGQEVARNRITLLSSGSFAVEEERVKYKD 360
      295 SLTAMTVEAMWGHAOSFPHLDVTRPPTQIEVAGNIRKLBSGRPAVHHEEVRKYKD 354
      361 EGILRODRYPLRTSPQFLGPIVEDMMAVSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 419
      355 EGILRODRYPLRTSPQFLGPIVEDMMAVSTLSLE--NNTTNDNPLDVENKOTAHGNFQ 414
      420 ASAVSISMKEKRLALAIIGKINFTQCTELLNANRGIPSCLAEDPSLHYGGLDHI 479
      415 AAIVANWMEKRLGIAQIGKINFTQCTELLNANRGIPSCLAEDPSLHYGGLDHI 474
      480 AAIVASIELGLANPVTTFFVQPAEMGNOAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
      475 AATSTELGLANPVTTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCYLOA 534
      540 VDLRAMELDFFKQFDPPLPTLLQOHG---TGLDV-NALALEVKKALKRLEQTTTYDLE 595
      535 IDLRAIEFEFFKQFGPAIVSLIDQHFSGAMTGSNLRDELVEKNKTLAKRLEQTNISYDLV 594
      596 PRMHDAFSYATGTVEYELSSPSANVTITAVANAMVSAEKAISLTREVRNRFQOTSSQ 655
      595 PRMHDAFSYATGTVEYELSSPSANVTITAVANAMVSAEKAISLTREVRNRFQOTSSQ 651
      656 APAAVYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHV 715
      652 SPALSTYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHV 711
      716 VKMLA 720
      712 LKMLA 716

RESULT 6
US-10-138-970A-4
; Sequence 4, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sarsaiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; TITLE OF INVENTION: hydroxycinnamic Acid
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; PRIOR FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: C. violaceum
US-10-138-970A-4

Query Match      72.4%; Score 2614; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-229;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1 MAPSLDSLATTLANGFTNGSHAAPTSAAGPTSAIRRTPTGLDGHAAHOSQLEIYQELISD 60
      1 MAPSLDSISHSFANG-----VASAKQAVNGASTWLAAGSHLPTTQVTOYDIVERKMLAA 54
      61 PTDDVVELSGYSLTVADVGAARKGRVRVONDEIRARVDKSVDFLKAQONSYGYGTT 120
      55 PTSTLELDGYSLNIGDVVSAARKGRPVAVKSDIRSKIDKSVDFLKAQONSYGYGTT 114
      121 GFGGSADRTTEDAVSLQKALIEHOLCGVTPPSXSFSGRGLENTLPLEVVRGAMVIRVN 180

```

Db 115 GFGSADRTEDAI SLQKALLEHQCGVLPSSGDSRLGRGLNSLPLEVVRGANTIRVN 174
Qy 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDSLPSLYAGAITGHPDVKYH 240
Db 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGTISASGDSLPSLYAGAITGHPDVKYH 234
Qy 241 VHEGTEKIMFARBAISLFGLEAVYLGPKEGLGVNGTAVSASATLSLHDSHMLSLISQ 300
Db 235 VHEGTEKIMFARBAISLFGLEAVYLGPKEGLGVNGTAVSASATLSLHDSHMLSLISQ 294
Qy 301 ALTALTEAMVGGQSPAPFIHDVCRPHGQVEVARNITRLSGSPFAVEHEEYKVKD 360
Db 295 SLTMTVEAMVGHAGSHRPHLDVTRPHPTQIEVAGNIRKLESGRPAVHHEEYKVKD 354
Qy 361 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 419
Db 355 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 414
Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNANRGLPSCLAEDPSLHYHKGKLDIHI 479
Db 415 ASAVSISMEKTRIALALIGKLNFTQCTELNANRGLPSCLAEDPSLHYHKGKLDIHI 474
Qy 480 AAYASELGHANPVTTFVQPAEMGNQAVNSIALISARTEANDVLSLLASHLYCTLOA 539
Db 475 AAYASELGHANPVTTFVQPAEMGNQAVNSIALISARTEANDVLSLLASHLYCTLOA 534
Qy 540 VDLRAMELDPKQOPDEPLTLLOQHIG---TGLDV-NALALEYKALKRELEQTTYYDLE 595
Db 535 IDLRALIEFEKQKQFAGLIVSLIDQHFSGAMTGSNRLDELVEKNTLAKRELEQTTYYDLE 594
Qy 596 PRMHDAFSYATGTVEVLELSSPSANVTLTAVNAKVASAKAISLTREVRNRFMQTPSSQ 655
Db 595 PRMHDAFSYATGTVEVLELSSPSANVTLTAVNAKVASAKAISLTREVRNRFMQTPSSQ 651
Qy 656 APAAAYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKGRINHYL 715
Db 652 SPALSYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKGRINHYL 711
Qy 716 VKMLA 720
Db 712 LKMLA 716

RESULT 7
US-10-188-523B-8
; Sequence 8, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; PRIOR FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodocorula glutinis
US-10-188-523B-8

Query Match 72.4%; Score 2614; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 9,5e-223;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

Qy 1 MAPSLDSTATTLANGFTNGSHAAPTSAAGPTSAIRTPGLDGHAAHQSCLEIVQOELSLD 60
Db 1 MAPSLDSTATTLANGFTNGSHAAPTSAAGPTSAIRTPGLDGHAAHQSCLEIVQOELSLD 54

Qy 61 PTDDVLEISGYSLITVRDVYGAARKGRVRYONDEIRARVDKSYDFKAOLONSVYGVTT 120
Db 55 PTDDVLEISGYSLITVRDVYGAARKGRVRYONDEIRARVDKSYDFKAOLONSVYGVTT 114
Qy 121 GFGSADRTEDAI SLQKALLEHQCGVLPSSGDSRLGRGLNSLPLEVVRGANTIRVN 180
Db 115 GFGSADRTEDAI SLQKALLEHQCGVLPSSGDSRLGRGLNSLPLEVVRGANTIRVN 174
Qy 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISASGDSLPSLYAGAITGHPDVKYH 240
Db 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGTISASGDSLPSLYAGAITGHPDVKYH 234
Qy 241 VHEGTEKIMFARBAISLFGLEAVYLGPKEGLGVNGTAVSASATLSLHDSHMLSLISQ 300
Db 235 VHEGTEKIMFARBAISLFGLEAVYLGPKEGLGVNGTAVSASATLSLHDSHMLSLISQ 294
Qy 301 ALTALTEAMVGGQSPAPFIHDVCRPHGQVEVARNITRLSGSPFAVEHEEYKVKD 360
Db 295 SLTMTVEAMVGHAGSHRPHLDVTRPHPTQIEVAGNIRKLESGRPAVHHEEYKVKD 354
Qy 361 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 419
Db 355 EGIIRQDRYPLRTSPQWLGPLVSDLIHAHVAVLTEAGQSTTDNPLIDVENKTSHHGQNFQ 414
Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNANRGLPSCLAEDPSLHYHKGKLDIHI 479
Db 415 ASAVSISMEKTRIALALIGKLNFTQCTELNANRGLPSCLAEDPSLHYHKGKLDIHI 474
Qy 480 AAYASELGHANPVTTFVQPAEMGNQAVNSIALISARTEANDVLSLLASHLYCTLOA 539
Db 475 AAYASELGHANPVTTFVQPAEMGNQAVNSIALISARTEANDVLSLLASHLYCTLOA 534
Qy 540 VDLRAMELDPKQOPDEPLTLLOQHIG---TGLDV-NALALEYKALKRELEQTTYYDLE 595
Db 535 IDLRALIEFEKQKQFAGLIVSLIDQHFSGAMTGSNRLDELVEKNTLAKRELEQTTYYDLE 594
Qy 596 PRMHDAFSYATGTVEVLELSSPSANVTLTAVNAKVASAKAISLTREVRNRFMQTPSSQ 655
Db 595 PRMHDAFSYATGTVEVLELSSPSANVTLTAVNAKVASAKAISLTREVRNRFMQTPSSQ 651
Qy 656 APAAAYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKGRINHYL 715
Db 652 SPALSYLSPTRTVLVSFVREBELGVQARRGDVFGVQOETTIGSNVRIYEAIKGRINHYL 711
Qy 716 VKMLA 720
Db 712 LKMLA 716

RESULT 8
US-10-374-366-2
; Sequence 2, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CII794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; PRIOR FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-2

Query Match 72.4%; Score 2614; DB 15; Length 716;
Best Local Similarity 72.0%; Pred. No. 9,5e-223;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

PRIOR APPLICATION NUMBER: US 09/627,216
 PRIOR FILING DATE: 2000-07-27
 PRIOR APPLICATION NUMBER: US 60/147,719
 PRIOR FILING DATE: 1999-08-06
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: Microsoft Office 97
 SEQ ID NO: 35
 LENGTH: 716
 TYPE: PRT
 ORGANISM: mutant from Rhodotorula glutinis
 FEATURE:
 NAME/KEY: UNSURE
 LOCATION: (198)..(198)
 OTHER INFORMATION: X= Asp, Asn, Glu, Gln
 NAME/KEY: UNSURE
 LOCATION: (540)..(540)
 OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
 US-09-765-873a-35

Query Match 72.4%; Score 2613; DB 9; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1,2e-228;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSATTLTANGFTNGSHAPRTSAAGPTSAURRTPGLDGHAHQSOLEIYQELISD 60
 DB 1 MAPSLDLSHSFANG-----VASAKQAVNGASTMLAVAGSHLPTQVTDIVEXMLAA 54
 QY 61 PTDVVELSGYSTVADVGAARKGRVAVNDDEIRARVDSVDFLAQIQNSVYGVTT 120
 DB 55 PTSTLELDGYSNLGDVVAARKGRPVVKDSDEIRSKIDKSEVFLSSQLSMSVYGVTT 114
 QY 121 GFGSADRTEDAVSLQKALIEHQCGVPTXSFSVSGLENTLPLEVVGAMVIRVN 180
 DB 115 GFGSADRTEDAVSLQKALIEHQCGVPTXSFSVSGLENTLPLEVVGAMVIRVN 174
 QY 181 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDSLSTIAGATGHPDKVH 240
 DB 175 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDSLSTIAGATGHPDKVH 234
 QY 241 VLHGTEKIMFARBAISLFGLEAVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
 DB 235 VVHGKEKILYARBAALFNLEPVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 294
 QY 301 ALTALTYEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
 DB 295 SLTAMTYEAMVGHAGSFHFDVTRPHPTQIEVAGNIRKLEGSRAVHHEEVKXKD 354
 QY 361 EGLLRDRYPLRTSPQFLGPLEVDMGAVSTLSL-NTTNDNPLDVENKQTAHGNFQ 419
 DB 355 EGLLRDRYPLRTSPQFLGPLEVDMGAVSTLSL-NTTNDNPLDVENKQTAHGNFQ 414
 QY 420 ASAVSISMEKRLALALIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 479
 DB 415 AAVALNTYKTRGLAIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 474
 QY 480 AAVALNTYKTRGLAIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 479
 DB 475 AAVALNTYKTRGLAIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 474
 QY 540 VDLRAVELDFKKQDFLPTLLOHNG---TGLDV-NALALEVVKALKNRLEQTTVDLE 595
 DB 535 IDLRARKEPFRKQFGALVSLIDQHFSGAMGNSRLDELVEKAKTKLAKLEQNSVDLV 594
 QY 596 PRNHDAFSYATGTVELLSSPSANVTLTAVANMVASAEKASLTREVRNRFWQTPSSQ 655
 DB 595 PRNHDAFSYATGTVELLSSPSANVTLTAVANMVASAEKASLTREVRNRFWQTPSSQ 651
 QY 656 APAAHATLSPRTVLVSFVVEELGVQARGVDVFGVQOCTIGSNVSRITYEALIKORINHL 715
 DB 652 SPALSTLSPRTQILAFVVEELGVQARGVDVFLGQEVTLIGSNVSKITYEALIKORINHL 711
 QY 716 VKMLA 720
 :|||

DB 712 LKMLA 716

RESULT 11
 US-10-138-970A-19
 Sequence 19, Application US/10138970A
 Publication No. US20030079255A1
 GENERAL INFORMATION:
 APPLICANT: Satiastani, Sima
 APPLICANT: Tang, Xiao-Song
 APPLICANT: Qi, Wei Wei
 TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
 FILE REFERENCE: C11777
 CURRENT APPLICATION NUMBER: US/10/138, 970A
 NUMBER OF SEQ ID NOS: 24
 SOFTWARE: Microsoft Office 97
 SEQ ID NO: 19
 LENGTH: 716
 TYPE: PRT
 ORGANISM: Rhodotorula glutinis
 FEATURE:
 NAME/KEY: MISC FEATURE
 LOCATION: (198)..(198)
 OTHER INFORMATION: X=Asp, Asn, Glu, Gln
 NAME/KEY: MISC FEATURE
 LOCATION: (198)..(198)
 OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
 US-10-138-970A-19

Query Match 72.4%; Score 2613; DB 14; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1,2e-228;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSATTLTANGFTNGSHAPRTSAAGPTSAURRTPGLDGHAHQSOLEIYQELISD 60
 DB 1 MAPSLDLSHSFANG-----VASAKQAVNGASTMLAVAGSHLPTQVTDIVEXMLAA 54
 QY 61 PTDVVELSGYSTVADVGAARKGRVAVNDDEIRARVDSVDFLAQIQNSVYGVTT 120
 DB 55 PTSTLELDGYSNLGDVVAARKGRPVVKDSDEIRSKIDKSEVFLSSQLSMSVYGVTT 114
 QY 121 GFGSADRTEDAVSLQKALIEHQCGVPTXSFSVSGLENTLPLEVVGAMVIRVN 180
 DB 115 GFGSADRTEDAVSLQKALIEHQCGVPTXSFSVSGLENTLPLEVVGAMVIRVN 174
 QY 181 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDSLSTIAGATGHPDKVH 240
 DB 175 SLTRGSHAVRLVLEALTNFLNHRITPIVLRGSIASGDSLSTIAGATGHPDKVH 234
 QY 241 VLHGTEKIMFARBAISLFGLEAVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
 DB 235 VVHGKEKILYARBAALFNLEPVVIGPKGGLGVNGTAVASAMATLSLHSHMLSLSQ 294
 QY 301 ALTALTYEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
 DB 295 SLTAMTYEAMVGHAGSFHFDVTRPHPTQIEVAGNIRKLEGSRAVHHEEVKXKD 354
 QY 361 EGLLRDRYPLRTSPQFLGPLEVDMGAVSTLSL-NTTNDNPLDVENKQTAHGNFQ 419
 DB 355 EGLLRDRYPLRTSPQFLGPLEVDMGAVSTLSL-NTTNDNPLDVENKQTAHGNFQ 414
 QY 420 ASAVSISMEKRLALALIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 479
 DB 415 AAVALNTYKTRGLAIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 474
 QY 480 AAVALNTYKTRGLAIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 479
 DB 475 AAVALNTYKTRGLAIGKLNFTQCTELLNANMNGRLPSCLAAEDPSLNYHGKGLDHI 474
 QY 540 VDLRAVELDFKKQDFLPTLLOHNG---TGLDV-NALALEVVKALKNRLEQTTVDLE 595
 DB 535 IDLRARKEPFRKQFGALVSLIDQHFSGAMGNSRLDELVEKAKTKLAKLEQNSVDLV 594
 QY 596 PRNHDAFSYATGTVELLSSPSANVTLTAVANMVASAEKASLTREVRNRFWQTPSSQ 655
 DB 595 PRNHDAFSYATGTVELLSSPSANVTLTAVANMVASAEKASLTREVRNRFWQTPSSQ 651
 QY 656 APAAHATLSPRTVLVSFVVEELGVQARGVDVFGVQOCTIGSNVSRITYEALIKORINHL 715
 DB 652 SPALSTLSPRTQILAFVVEELGVQARGVDVFLGQEVTLIGSNVSKITYEALIKORINHL 711
 QY 716 VKMLA 720
 :|||

```

Db 475 AAYTSELGHLANPVTTHQPAEMANQAVNSIALISARRTESNDVLSLLATLHYCYLQA 534
QY 540 VDIRAMEIDFKKQFPLPPTLLOQHLC---TGLDV-NALALEYKALKNELEOTTYDLE 595
Db 535 IDLRAXEPEFKQFQGPALVSLIDQFSGAMTGSNLRBELVEKNKTLAKLEOTNSYDLV 594
QY 596 PRMHDAFSGATGVVELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 655
Db 595 PRMHDAFSGATGVVELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 651
QY 656 APAAHYLSPTRTVLVYSPFREELGVQARGDVYVGVQOETIGSNVSRLYEAIKQGRINHL 715
Db 652 SPALSYLSPTRTQILYAFVREELGVQARGDVFLGKQEVLTIGSNVSKTYEAIKSGRINNVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

```

```

RESULT 12
US-10-188-523B-35
; Sequence 35, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVICI
; CURRENT APPLICATION NUMBER: US/10/188,523B
; PRIOR FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 35
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis mutant
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa= Asp, Asn, Glu, or Gln
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa= Thr, Ala, Ser, Pro, or Gly
US-10-188-523B-35

```

```

Query Match 72.4%; Score 2613; DB 14; Length 716;
Best Local Similarity 72.0%; Pred. No. 1,2e-228;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLSLTTTANGTNGSHAAFTSKAAPTSAIRRTGDLGHAHQSOLEIVQELSD 60
Db 1 MAPSLSLSHSFGNG-----VASAKQAVNGASTLVAAGSLPTTYQTVQDYIEKMLAA 54
QY 61 PTDVVELSGYSLTVADVGAARKGRVAVQNDDEIRAVDKSVDFLPAQLQNSYGVTT 120
Db 55 PTDSTLELDGYSLNLDGVYSAARKGRVAVKDSDEIRKSDXSVFELRSQLSMXYGVTT 114
QY 121 GFGGSADTRIEDAVSQXLIIEHQLCGVPTKXSPFVGRGLENTLPLEVVRGAMTVRN 180
Db 115 GFGGSADTRIEDAISIQLKALEHQLCGVLPSSFDSPRLGRGLENSLPLEVVRGAMTVRN 174
QY 181 SLTRGSAVRLVLEALTNFLNRRITPIVPLRGSSISAGSDSPSYIAGATGHPDVVH 240
Db 175 SLTRGSAVRLVLEALTNFLNRRITPIVPLRGSSISAGSDSPSYIAGATGHPDVVH 234
QY 241 VHEGTEKTMFAPEAISIQLGLENVTVGPEKGIQLVNGTVKVSAMTTLHLHSHMISLSIQ 300
Db 235 VHEGTEKTLIYAREAMALFNLPEVVLGPRKGLGVNGLTAVASAMTTLALHDAHMLISLSIQ 294

```

```

QY 301 ALTAITVAMVGOQGSFAFTHDYCRPHQGVARNIRTLSSGSPAVHEBEVKYKD 360
Db 295 SLIMTAVAMVGHAGSHPHFLHDVTRPHPTQIEVAGNIRKLBSGRPAVHHEBEVKYKD 354
QY 361 EGI LRQDRIYPIRTSPQFLGPIVEMDMAYSTLSIE-NNTTNDPLDVEKQRAHGNFQ 419
Db 355 EGI LRQDRIYPIRTSPQFLGPIVEMDMAYSTLSIE-NNTTNDPLDVEKQRAHGNFQ 414
QY 420 ASAVSISNEKTRLMALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGGLDIHI 479
Db 415 AAAVANTHEKTRLGLAQGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGGLDIHI 474
QY 480 AAYASELGHANPVTTHQPAEMANQAVNSIALISARRTESNDVLSLLATLHYCYLQA 534
Db 475 AAYTSELGHLANPVTTHQPAEMANQAVNSIALISARRTESNDVLSLLATLHYCYLQA 534
QY 540 VDIRAMEIDFKKQFPLPPTLLOQHLC---TGLDV-NALALEYKALKNELEOTTYDLE 595
Db 535 IDLRAXEPEFKQFQGPALVSLIDQFSGAMTGSNLRBELVEKNKTLAKLEOTNSYDLV 594
QY 596 PRMHDAFSGATGVVELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 655
Db 595 PRMHDAFSGATGVVELLSSPSANVTLTAVANAKVSAEKAISLTREVENRFWQTPSSQ 651
QY 656 APAAHYLSPTRTVLVYSPFREELGVQARGDVYVGVQOETIGSNVSRLYEAIKQGRINHL 715
Db 652 SPALSYLSPTRTQILYAFVREELGVQARGDVFLGKQEVLTIGSNVSKTYEAIKSGRINNVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

```

```

RESULT 13
US-10-439-478-15
; Sequence 15, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariassani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Vansili, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL1912
; CURRENT APPLICATION NUMBER: US/10/439,478
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 15
; LENGTH: 716
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: Xaa= Asp, Asn, Glu, or Gln
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa= Thr, Ala, Ser, Pro, or Gly
US-10-439-478-15

```

```

Query Match 72.4%; Score 2613; DB 15; Length 716;
Best Local Similarity 72.0%; Pred. No. 1,2e-228;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```

```

QY 1 MAPSLDLSLATTIANGFTNGSHAAPTKSAAGPTALRTTGLDGHAAHQSLTVOGLSLSD 60
DB 1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDVVELSGYSLTRVDVGAARKGRVRVONDEIRARVKSVDPLKAOQNSVYGVTT 120
DB 55 PTDSTLELDGYSNLNGDVSAARKGRVPRVKDSDEIRSKIDKSVEILRSQLSMSYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGYTPTSXSSFSVGRGLENTLPLEVVGAMVIRVN 180
DB 115 GFGGSADRTEDAVSLQKALIEHQLCGYTPTSXSSFSVGRGLENTLPLEVVGAMVIRVN 174
QY 181 SLTRGSAVRVLVLEALTNFLNRTTPVLRGSIASGDSLPLSTAGATIGHGPDVYKH 240
DB 175 SLTRGSAVRVLVLEALTNFLNRTTPVLRGSIASGDSLPLSTAGATIGHGPDVYKH 234
QY 241 VLHGTGKIMFARBAISLFGLEAVVLGPKREGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
DB 235 VLHGTGKIMFARBAISLFGLEAVVLGPKREGGLGVNGTAVASAMATLSLHSHMLSLSQ 294
QY 301 ALTALTVEAMVGGQGSFAPFIHDVCRPHGQVEVARNIRTLSSGSPFAVHEEVEVKYKD 360
DB 295 SLTMTVEAMVGGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLEGRFPAVHEEVEVKYKD 354
QY 361 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHAGNFQ 419
DB 355 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHAGNFQ 414
QY 420 ASAVSISEKTRIALALIGKLTNCTQTELLNAAMNRLPSCLAEDPSLNYHKGGLDIH 479
DB 415 AAIVANMTEKTRIGLAIQIKLNTFTQLEMLNAGMNRGLPSCLAEDPSLSYHKGGLDIAA 474
QY 480 AAIVASELGHANPVTTPVOPAEKMGNOAVNSLALISARTEANDVSLILASLHLYCTLOA 539
DB 475 AAIVASELGHANPVTTPVOPAEKMGNOAVNSLALISARTEANDVSLILASLHLYCTLOA 534
QY 540 VDIRAMELDJFKQFDPPLPTLLQOHLG---TGLDV-NALALEVKAALKRLEQTTVDLE 595
DB 535 IDIRAFEFEPKQFGAIVSLIDQHFSGAMTGSNLRDELVEKNTLAKRLEQNTSDLV 594
QY 596 PRMHDAFSYATGTVEVLLSSPSANVTLLAVNAMKVASAKAISLITREVRNRWQTPSSQ 655
DB 595 PRMHDAFSYATGTVEVLLSSPSANVTLLAVNAMKVASAKAISLITREVRNRWQTPSSQ 651
QY 656 APAHAYLSPTREVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHYL 715
DB 652 SPALSTYLSPTQIILYAFVREELGVKARRGDVFLGKQEVITIGSNVSKIYEAIKGRINHYL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

```

RESULT 14

```

US-09-765-873a-10
/ Sequence 10, Application US/09765873A
/ Patent No. US20010053847A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Xiao-Song
/ TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
/ FILE REFERENCE: BC1009 US CIP
/ CURRENT APPLICATION NUMBER: US/09/765, 873A
/ PRIOR FILING DATE: 2001-01-19
/ PRIOR APPLICATION NUMBER: US 09/627, 216
/ PRIOR FILING DATE: 2000-07-27
/ PRIOR APPLICATION NUMBER: US 60/147, 719
/ PRIOR FILING DATE: 1999-08-06
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 10
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: mutant from Rhodotorula glutinis

```

US-09-765-873a-10

Query Match 72.4%; Score 2612; DB 9; Length 716;
 Best Local Similarity 72.0%; Pred. No. 1,4e-228;
 Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```

QY 1 MAPSLDLSLATTIANGFTNGSHAAPTKSAAGPTALRTTGLDGHAAHQSLTVOGLSLSD 60
DB 1 MAPSLDLSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PTDVVELSGYSLTRVDVGAARKGRVRVONDEIRARVKSVDPLKAOQNSVYGVTT 120
DB 55 PTDSTLELDGYSNLNGDVSAARKGRVPRVKDSDEIRSKIDKSVEILRSQLSMSYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALIEHQLCGYTPTSXSSFSVGRGLENTLPLEVVGAMVIRVN 180
DB 115 GFGGSADRTEDAVSLQKALIEHQLCGYTPTSXSSFSVGRGLENTLPLEVVGAMVIRVN 174
QY 181 SLTRGSAVRVLVLEALTNFLNRTTPVLRGSIASGDSLPLSTAGATIGHGPDVYKH 240
DB 175 SLTRGSAVRVLVLEALTNFLNRTTPVLRGSIASGDSLPLSTAGATIGHGPDVYKH 234
QY 241 VLHGTGKIMFARBAISLFGLEAVVLGPKREGGLGVNGTAVASAMATLSLHSHMLSLSQ 300
DB 235 VLHGTGKIMFARBAISLFGLEAVVLGPKREGGLGVNGTAVASAMATLSLHSHMLSLSQ 294
QY 301 ALTALTVEAMVGGQGSFAPFIHDVCRPHGQVEVARNIRTLSSGSPFAVHEEVEVKYKD 360
DB 295 SLTMTVEAMVGGHAGSFHPFLHDVTRPHPTQIEVAGNIRKLEGRFPAVHEEVEVKYKD 354
QY 361 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHAGNFQ 419
DB 355 EGLRQDRYPLRTSPQFLGPLEVDMMAVSTLSLE-NTTNDNPLDVENKQTAHAGNFQ 414
QY 420 ASAVSISEKTRIALALIGKLTNCTQTELLNAAMNRLPSCLAEDPSLNYHKGGLDIH 479
DB 415 AAIVANMTEKTRIGLAIQIKLNTFTQLEMLNAGMNRGLPSCLAEDPSLSYHKGGLDIAA 474
QY 480 AAIVASELGHANPVTTPVOPAEKMGNOAVNSLALISARTEANDVSLILASLHLYCTLOA 539
DB 475 AAIVASELGHANPVTTPVOPAEKMGNOAVNSLALISARTEANDVSLILASLHLYCTLOA 534
QY 540 VDIRAMELDJFKQFDPPLPTLLQOHLG---TGLDV-NALALEVKAALKRLEQTTVDLE 595
DB 535 IDIRAFEFEPKQFGAIVSLIDQHFSGAMTGSNLRDELVEKNTLAKRLEQNTSDLV 594
QY 596 PRMHDAFSYATGTVEVLLSSPSANVTLLAVNAMKVASAKAISLITREVRNRWQTPSSQ 655
DB 595 PRMHDAFSYATGTVEVLLSSPSANVTLLAVNAMKVASAKAISLITREVRNRWQTPSSQ 651
QY 656 APAHAYLSPTREVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHYL 715
DB 652 SPALSTYLSPTQIILYAFVREELGVKARRGDVFLGKQEVITIGSNVSKIYEAIKGRINHYL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

```

RESULT 15

```

US-10-138-970a-24
/ Sequence 24, Application US/10138970A
/ Publication No. US20030079255A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Xiao-Song
/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ FILE REFERENCE: C11777
/ CURRENT APPLICATION NUMBER: US/10/138, 970A
/ PRIOR FILING DATE: 2002-07-23
/ NUMBER OF SEQ ID NOS: 24

```

```
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 24
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-138-970A-24
```

Query Match 72.4%; Score 2612; DB 14; Length 716;

Best Local Similarity 72.0%; Pred. No. 1,4e-228; Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```
1 MAPSLDSLATTANGFTNGSHAAPTSAAGPTSAALRRTPGLDGHAHQSOLEIVQELSLD 60
1 MAPSLDSISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVYDIKMLAA 54
61 PTDVVELSGYGLTVRDVVGAAKRGRRVQNDDEIRARVDKSVDFLKAQQLQNSVYGVTT 120
55 PTDSTLELDGSLNGDVVSAARKGRPVYKDSDEIRSKIDKSVDFLRSLMSVYGVTT 114
121 GFGSGADRTREDVAVSLQKALIEHOLCGVTPTSKSSFSVGRGIENTLPLEVYRGAMVIRVN 180
115 GFGSGADRTEDVAISIQKALIEHOLCGVLPSEFDSFRLGRGLNSLPLEVYRGAMTIRVN 174
181 SLTRGSAVRLVYLALTNFLNHRITPIVPLRGSISASGDLSPSYIAGAITGHPDYKH 240
175 SLTRGSAVRLVYLALTNFLNHRITPIVPLRGRISASGDLSPSYIAAISGHPDSKH 234
241 VHEGTEKTMFARALSLFGLAAVVLGREGGLVNGTAVASMATLSLHSHMLSLSQ 300
235 VHEGTEKTLVAREKALFNLFPVVLGREGGLVNGTAVASMATLALHDAHMLSLSQ 294
301 ALTALTVEAMVQCGSPAFFIDVCRPHPGVEVARNIRITLLSGSSFAVEHEEEVKYKD 360
295 SLTAMTVEAMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKLEGSRFVHHEEEVKYKD 354
361 EGLLRQDRYPLRTSPQFLGPIVEDMMAHAYSTLSLE-NNTTDNPLLDYENKQTAHGNFQ 419
355 EGLLRQDRYPLRTSPQFLGPIVEDMMAHAYSTLSLE-NNTTDNPLLDYENKSHHGNFQ 414
420 ASAVSISMEKTLALALIGKLNFTQCTELNAAMNRGLPSCLAADPSLNYHKGDLHI 479
415 AAAVANTMEKTLGLAQIGKLNFTQCTELNAAMNRGLPSCLAADPSLNYHKGDLIAA 474
480 AAAYASELGHANPVTTFFVQPAEMGNQAVNSLALISARRTAANDVLSLLASHLYCTLOA 539
475 AAYTSELGHANPVTTTHVQPAEMANOAVNSLALISARRTESNDVLSLLATHLYCVLOA 534
540 VDIRAMELDFKQDFDLPTLLQOHG--TGLDV-NALALEVKKALKRLQOTTYYDLE 595
535 IDLRATEFEFFKQDFGPAIVSLIDQHFSGAMTGSNLDLVEKVKNTLAKRLEQTNSTYDLV 594
596 PRWHDASFATGIVVVELSSPSANVTLLAVAMKVASAEKASITREVRNRFWQTPSSQ 655
595 PRWHDASFATGIVVVELSSPSANVTLLAVAMKVASAEKASITREVRNRFWQTPSSQ 651
656 APAAHAYLSPTRYLVSFVEEELGVQARRQDVFGVQCEITIGSNVRIYEAIKQGRINHL 715
652 SPALSTYLSPTQIOLYAFVEEELGVKARRQDVFLGKQEVITIGSNVSKIYEAIKSGRINVL 711
716 VKMLA 720
712 LKMLA 716
```

Search completed: September 9, 2004, 10:24:24
Job time : 95.6141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:04:14 ; Search time 295.768 Seconds

(without alignments)
2376.050 Million cell updates/sec

Title: US-09-939-408A-13

Perfect score: 3610
Sequence: 1 MAPSIDSATLTIANGFTNGS.....RIYRAIDGRINHYLMQLA 720

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA Main:*

1:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*
2:	/cgn2_6/ptodata/2/paa/US06_COMB.pep.*
3:	/cgn2_6/ptodata/2/paa/US07_COMB.pep.*
4:	/cgn2_6/ptodata/2/paa/US08_COMB.pep.*
5:	/cgn2_6/ptodata/2/paa/US081_COMB.pep.*
6:	/cgn2_6/ptodata/2/paa/US082_COMB.pep.*
7:	/cgn2_6/ptodata/2/paa/US083_COMB.pep.*
8:	/cgn2_6/ptodata/2/paa/US084_COMB.pep.*
9:	/cgn2_6/ptodata/2/paa/US085_COMB.pep.*
10:	/cgn2_6/ptodata/2/paa/US086_COMB.pep.*
11:	/cgn2_6/ptodata/2/paa/US087_COMB.pep.*
12:	/cgn2_6/ptodata/2/paa/US088_COMB.pep.*
13:	/cgn2_6/ptodata/2/paa/US089_COMB.pep.*
14:	/cgn2_6/ptodata/2/paa/US090_COMB.pep.*
15:	/cgn2_6/ptodata/2/paa/US091_COMB.pep.*
16:	/cgn2_6/ptodata/2/paa/US092_COMB.pep.*
17:	/cgn2_6/ptodata/2/paa/US093_COMB.pep.*
18:	/cgn2_6/ptodata/2/paa/US094_COMB.pep.*
19:	/cgn2_6/ptodata/2/paa/US095_COMB.pep.*
20:	/cgn2_6/ptodata/2/paa/US096_COMB.pep.*
21:	/cgn2_6/ptodata/2/paa/US097_COMB.pep.*
22:	/cgn2_6/ptodata/2/paa/US097A_COMB.pep.*
23:	/cgn2_6/ptodata/2/paa/US098_COMB.pep.*
24:	/cgn2_6/ptodata/2/paa/US098A_COMB.pep.*
25:	/cgn2_6/ptodata/2/paa/US099_COMB.pep.*
26:	/cgn2_6/ptodata/2/paa/US099A_COMB.pep.*
27:	/cgn2_6/ptodata/2/paa/US101_COMB.pep.*
28:	/cgn2_6/ptodata/2/paa/US102_COMB.pep.*
29:	/cgn2_6/ptodata/2/paa/US103_COMB.pep.*
30:	/cgn2_6/ptodata/2/paa/US104_COMB.pep.*
31:	/cgn2_6/ptodata/2/paa/US106_COMB.pep.*
32:	/cgn2_6/ptodata/2/paa/US107_COMB.pep.*
33:	/cgn2_6/ptodata/2/paa/US00_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	3608	99.9	720	1	PCT-US02-26934A-13	Sequence 13, Appl
2	3608	99.9	720	24	US-09-939-408A-13	Sequence 13, Appl
3	2770	76.7	726	1	PCT-US02-26934A-21	Sequence 21, Appl
4	2770	76.7	726	24	US-09-939-408A-21	Sequence 21, Appl
5	2730	75.6	720	1	PCT-US02-26934A-30	Sequence 30, Appl
6	2730	75.6	720	24	US-09-939-408A-30	Sequence 30, Appl
7	2614	72.4	716	1	PCT-US02-26934A-19	Sequence 19, Appl
8	2614	72.4	716	1	PCT-US03-05708-2	Sequence 2, Appl
9	2614	72.4	716	1	PCT-US03-17926-1	Sequence 2, Appl
10	2614	72.4	716	24	US-09-939-408A-19	Sequence 19, Appl
11	2614	72.4	716	27	US-10-188-970A-4	Sequence 8, Appl
12	2614	72.4	716	27	US-10-188-523C-8	Sequence 8, Appl
13	2614	72.4	716	27	US-10-188-523C-8	Sequence 2, Appl
14	2614	72.4	716	29	US-10-374-366-2	Sequence 2, Appl
15	2614	72.4	716	30	US-60-147-719-8	Sequence 2, Appl
16	2614	72.4	716	33	US-60-147-719-8	Sequence 8, Appl
17	2614	72.4	716	33	US-60-383-089-4	Sequence 4, Appl
18	2614	72.4	716	33	US-60-460-120-4	Sequence 4, Appl
19	2613	72.4	716	1	PCT-US03-17926-15	Sequence 15, Appl
20	2613	72.4	716	27	US-10-138-970A-19	Sequence 19, Appl
21	2613	72.4	716	27	US-10-188-523B-15	Sequence 35, Appl
22	2613	72.4	716	27	US-10-188-523C-35	Sequence 35, Appl
23	2613	72.4	716	30	US-10-439-478-15	Sequence 15, Appl
24	2612	72.4	716	1	PCT-US03-05708-4	Sequence 4, Appl
25	2612	72.4	716	27	US-10-138-970A-24	Sequence 24, Appl
26	2612	72.4	716	27	US-10-188-523B-10	Sequence 10, Appl
27	2612	72.4	716	29	US-10-188-523C-10	Sequence 10, Appl
28	2612	72.4	716	29	US-60-147-719-10	Sequence 4, Appl
29	2612	72.4	716	33	US-60-147-719-10	Sequence 10, Appl
30	2609	72.3	716	29	PCT-US03-05708-24	Sequence 24, Appl
31	2609	72.3	716	29	US-10-374-366-24	Sequence 24, Appl
32	2608	72.2	716	1	PCT-US03-05708-22	Sequence 22, Appl
33	2608	72.2	716	29	US-10-374-366-22	Sequence 22, Appl
34	2607	72.2	716	1	PCT-US03-05708-16	Sequence 16, Appl
35	2607	72.2	716	1	PCT-US03-17926-16	Sequence 16, Appl
36	2607	72.2	716	27	US-10-138-970A-22	Sequence 22, Appl
37	2607	72.2	716	27	US-10-188-523B-38	Sequence 38, Appl
38	2607	72.2	716	27	US-10-188-523C-38	Sequence 38, Appl
39	2607	72.2	716	29	US-10-374-366-16	Sequence 16, Appl
40	2607	72.2	716	30	US-10-439-478-18	Sequence 18, Appl
41	2606	72.2	716	1	PCT-US03-05708-8	Sequence 8, Appl
42	2606	72.2	716	29	US-10-374-366-8	Sequence 8, Appl
43	2605	72.2	716	1	PCT-US03-05708-10	Sequence 10, Appl
44	2605	72.2	716	29	US-10-374-366-10	Sequence 10, Appl
45	2604	72.1	716	1	PCT-US03-05708-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
PCT-US02-26934A-13
Sequence 13, Application PC/TUS0226934A
GENERAL INFORMATION:
APPLICANT: PCBU Services, Inc. et al.
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
TITLE OF INVENTION: Using Same
FILE REFERENCE: 29479/50USCA PCT
CURRENT APPLICATION NUMBER: PCT/US02/26934A
CURRENT FILING DATE: 2002-08-23
PRIOR APPLICATION NUMBER: US 09/939,408
PRIOR FILING DATE: 2001-08-24
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodococcus graminis
FEATURE:
NAME/KEY: unsure
LOCATION: (153)
OTHER INFORMATION: Other information: Xaa = Val or Ala

Mon Sep 13 10:31:10 2004

us-09-939-408a-13.rapm

Page 2

PCT-US02-26934A-13

Query Match 99.9%; Score 3608; DB 1; Length 720;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAPSLDLSLATTLLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIVQELLSD 60
D 1 MAPSLDLSLATTLLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIVQELLSD 60
QY 61 PTDDVVELSGSYSLTVRDVVGAAKRGRRVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
D 61 PTDDVVELSGSYSLTVRDVVGAAKRGRRVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
QY 121 GFSGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
D 121 GFSGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
QY 181 SLTRGSAVRVLYVEALTNFNLNHRITPVLPRGSIASGDLSPSTYAGAITGHPDKVH 240
D 181 SLTRGSAVRVLYVEALTNFNLNHRITPVLPRGSIASGDLSPSTYAGAITGHPDKVH 240
QY 241 VLHGEKIMFAREALSLFGLEAVVLPKRGGLGVNGTAVASAMATLSLHSHMLSLLSQ 300
D 241 VLHGEKIMFAREALSLFGLEAVVLPKRGGLGVNGTAVASAMATLSLHSHMLSLLSQ 300
QY 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
D 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
QY 361 EGIIRDORVPLRTSPQFPLVEDMMHAYSTLSLNNNTTNDPLDVENKOTAHGNGFOA 420
D 361 EGIIRDORVPLRTSPQFPLVEDMMHAYSTLSLNNNTTNDPLDVENKOTAHGNGFOA 420
QY 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMNGLPSCLAEDPSLNHGKGLDIIHA 480
D 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMNGLPSCLAEDPSLNHGKGLDIIHA 480
QY 481 AVASELGHLANPVTTPVQPAEMGNQAVNSIALISARRTAANDVLSLLASHLYCTLQAV 540
D 481 AVASELGHLANPVTTPVQPAEMGNQAVNSIALISARRTAANDVLSLLASHLYCTLQAV 540
QY 541 DLRAEMLDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEBRWD 600
D 541 DLRAEMLDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEBRWD 600
QY 601 AFSYATGTVVELLSSSPSANVTLLTAVNAKVASAKAISLREVRNRPQOTPSQAPAA 660
D 601 AFSYATGTVVELLSSSPSANVTLLTAVNAKVASAKAISLREVRNRPQOTPSQAPAA 660
QY 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRIVEAIKDGRIINHLYKMLA 720
D 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRIVEAIKDGRIINHLYKMLA 720
```

RESULT 2
US-09-939-408a-13
Sequence 13, Application US/09939408A
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA
CURRENT APPLICATION NUMBER: US/09/939,408A
PRIOR FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 13
LENGTH: 720
TYPE: PRT
ORGANISM: Rhodotorula graminis
FEATURE:
NAME/KEY: unsure
LOCATION: (153)
OTHER INFORMATION: Other information: Xaa = Val or Ala
US-09-939-408a-13

Query Match 99.9%; Score 3608; DB 24; Length 720;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MAPSLDLSLATTLLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIVQELLSD 60
D 1 MAPSLDLSLATTLLANGFTNGSHAAPTKSAAGPTSAARRTPGLDGHAAHQSOLEIVQELLSD 60
QY 61 PTDDVVELSGSYSLTVRDVVGAAKRGRRVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
D 61 PTDDVVELSGSYSLTVRDVVGAAKRGRRVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
QY 121 GFSGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
D 121 GFSGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
QY 181 SLTRGSAVRVLYVEALTNFNLNHRITPVLPRGSIASGDLSPSTYAGAITGHPDKVH 240
D 181 SLTRGSAVRVLYVEALTNFNLNHRITPVLPRGSIASGDLSPSTYAGAITGHPDKVH 240
QY 241 VLHGEKIMFAREALSLFGLEAVVLPKRGGLGVNGTAVASAMATLSLHSHMLSLLSQ 300
D 241 VLHGEKIMFAREALSLFGLEAVVLPKRGGLGVNGTAVASAMATLSLHSHMLSLLSQ 300
QY 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
D 301 ALTALIVEAMVGOQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
QY 361 EGIIRDORVPLRTSPQFPLVEDMMHAYSTLSLNNNTTNDPLDVENKOTAHGNGFOA 420
D 361 EGIIRDORVPLRTSPQFPLVEDMMHAYSTLSLNNNTTNDPLDVENKOTAHGNGFOA 420
QY 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMNGLPSCLAEDPSLNHGKGLDIIHA 480
D 421 SAVSISMEKTRIALALIGKLNFTQCTELNNAAMNGLPSCLAEDPSLNHGKGLDIIHA 480
QY 481 AVASELGHLANPVTTPVQPAEMGNQAVNSIALISARRTAANDVLSLLASHLYCTLQAV 540
D 481 AVASELGHLANPVTTPVQPAEMGNQAVNSIALISARRTAANDVLSLLASHLYCTLQAV 540
QY 541 DLRAEMLDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEBRWD 600
D 541 DLRAEMLDFKQFPLPTLLQOHLGTGLDVNALALEVKKALNKRLEQTTTYDLEBRWD 600
QY 601 AFSYATGTVVELLSSSPSANVTLLTAVNAKVASAKAISLREVRNRPQOTPSQAPAA 660
D 601 AFSYATGTVVELLSSSPSANVTLLTAVNAKVASAKAISLREVRNRPQOTPSQAPAA 660
QY 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRIVEAIKDGRIINHLYKMLA 720
D 661 YLSPTRTVLYSFVARELGVQARRGDVFGVQOETIGSNVSRIVEAIKDGRIINHLYKMLA 720
```

RESULT 3
PCT-US02-26934A-21
Sequence 21, Application PC/TUS0226934A
GENERAL INFORMATION:
APPLICANT: PCB Services, Inc. et al.
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA PCT
CURRENT APPLICATION NUMBER: PCT/US02/26934A

Db 541 AYLRLRMEEFFKQGFYPPXXXLLXQHFQXXXLXXXXXXXXXXKXKXKXKXLEDTNSYDL 6000

QY 595 EPPMHDAFSAIGTIVTELLSSBSANVTITAVANAKVASAEKASILFREVNRFWQTPSS 6544

Db 601 EPPMHDAFSAIGTIVTELLSSBSANVTITAVANAKVASAEKASILTXVXXFXWAPSS 6600

QY 655 QAENAHVLSERTVILSVFVEELGYQARRGDVFVCOEITGSNVSRIYEAIKDGRINHV 7144

Db 661 SSPALXYLSPTRTVLVSFVREELGYQARRGDVFLKQEVITIGSNVSRIYEAIKGRINXV 7200

QY 715 LVKQLA 7200

Db 721 LVKQLA 726

```

RESULT 5
PCT-US02-26934A-30
: Sequence 30, Application PC/TUS0226934A
: GENERAL INFORMATION:
: APPLICANT: PCBU Services, Inc. et al.
: TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
: TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
: TITLE OF INVENTION: Using Same
: FILE REFERENCE: 29479/50NSCA PCT
: CURRENT APPLICATION NUMBER: PCT/US02/26934A
: PRIOR FILING DATE: 2002-08-23
: PRIOR APPLICATION NUMBER: US 09/939,408
: PRIOR FILING DATE: 2001-08-24
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 30
: LENGTH: 720
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Preferred
: OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (5)..(5)
: OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (12)..(12)
: OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (16)..(16)
: OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (17)..(17)
: OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (19)..(19)
: OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (20)..(20)
: OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (25)..(25)
: OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (27)..(27)
: OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (28)..(28)

```

```

1 OTHER INFORMATION: The 'Xaa' at location 28 strands for Ala, or Pro.
2 FEATURE:
3 NAME/KEY: misc_feature
4 LOCATION: (34)..(34)
5 OTHER INFORMATION: The 'Xaa' at location 34 strands for Thr.
6 FEATURE:
7 NAME/KEY: misc_feature
8 LOCATION: (35)..(35)
9 OTHER INFORMATION: The 'Xaa' at location 36 strands for Arg, or Ser.
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: (39)..(39)
13 OTHER INFORMATION: The 'Xaa' at location 39 strands for Ala, Pro, or Ser.
14 FEATURE:
15 NAME/KEY: misc_feature
16 LOCATION: (40)..(40)
17 OTHER INFORMATION: The 'Xaa' at location 40 strands for Arg, Gly, or Trp.
18 FEATURE:
19 NAME/KEY: misc_feature
20 LOCATION: (48)..(48)
21 OTHER INFORMATION: The 'Xaa' at location 48 strands for Lys, Thr, Met, Glu, Ala,
22 OTHER INFORMATION: Val, Gln, Pro, or Leu.
23 FEATURE:
24 NAME/KEY: misc_feature
25 LOCATION: (54)..(54)
26 OTHER INFORMATION: The 'Xaa' at location 54 strands for Val.
27 FEATURE:
28 NAME/KEY: misc_feature
29 LOCATION: (55)..(55)
30 OTHER INFORMATION: The 'Xaa' at location 56 strands for Lys, Glu, or Gln.
31 FEATURE:
32 NAME/KEY: misc_feature
33 LOCATION: (65)..(65)
34 OTHER INFORMATION: The 'Xaa' at location 65 strands for Glu, Asp, or Val.
35 FEATURE:
36 NAME/KEY: misc_feature
37 LOCATION: (66)..(66)
38 OTHER INFORMATION: The 'Xaa' at location 66 strands for Ile, Val, or Leu.
39 FEATURE:
40 NAME/KEY: misc_feature
41 LOCATION: (76)..(76)
42 OTHER INFORMATION: The 'Xaa' at location 76 strands for Gly.
43 FEATURE:
44 NAME/KEY: misc_feature
45 LOCATION: (87)..(87)
46 OTHER INFORMATION: The 'Xaa' at location 87 strands for Thr, Pro, or Ser.
47 FEATURE:
48 NAME/KEY: misc_feature
49 LOCATION: (93)..(93)
50 OTHER INFORMATION: The 'Xaa' at location 93 strands for Asp, or Ala.
51 FEATURE:
52 NAME/KEY: misc_feature
53 LOCATION: (102)..(102)
54 OTHER INFORMATION: The 'Xaa' at location 102 strands for Lys, or Asn.
55 FEATURE:
56 NAME/KEY: misc_feature
57 LOCATION: (103)..(103)
58 OTHER INFORMATION: The 'Xaa' at location 103 strands for Arg, or Ser.
59 FEATURE:
60 NAME/KEY: misc_feature
61 LOCATION: (112)..(112)
62 OTHER INFORMATION: The 'Xaa' at location 112 strands for Asp, His, or Tyr.
63 FEATURE:
64 NAME/KEY: misc_feature
65 LOCATION: (114)..(114)
66 OTHER INFORMATION: The 'Xaa' at location 114 strands for Arg, or Ser.
67 FEATURE:
68 NAME/KEY: misc_feature
69 LOCATION: (117)..(117)

```

```

OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (148)..(148)
OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150)..(150)
OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (153)..(153)
OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (154)..(154)
OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (157)..(157)
OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (159)..(159)
OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (183)..(183)
OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (223)..(223)
OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (225)..(225)
OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (237)..(237)
OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala, Val.
OTHER INFORMATION: Cys, or Phe.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (239)..(239)
OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (241)..(241)
OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (242)..(242)
OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (246)..(246)
OTHER INFORMATION: The 'Xaa' at location 246 stands for Lys, Asn, or Thr.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (251)..(251)
OTHER INFORMATION: The 'Xaa' at location 251 stands for Tyr, Ser, or Phe.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (259)..(259)
OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (305)..(305)
OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
FEATURE:
NAME/KEY: misc_feature

```

```

Query Match      75.6%; Score 2730; DB 1.; Length 720;
Best Local Similarity 76.5%; Pred. No. 3,76-258;
Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

QY      1 MAPSLSLATTTANGFTNGSHAAPTSAAGTSAATRTPGLDGHAHQSLRFVCELSLD 60
DB      1 MARSXSINATSAANGKXNXXHAPKXXXTGATGSXLMXPKLAPPAQXQLDXEITLAD 60

QY      61 PTDDVVELSGSLTFEDVYGAAKGRFRVONDEIFARVDKSVDFLKQJLONSVTGTT 120
DB      61 PTDDXXELDGYLTLLIXDVGAARKGRXVRVQTQXBEIAKIDXXVEFLRXQLNXXVXXVT 120

QY      121 GFGSADPTEDPADVSXQALIFHQJCGTPTFSXSFSVGRGENTPLEVVGAMVIRN 180
DB      121 GFGSADPTEDPADISIQRLLEHQCGKLTXSXSPFLXGLGENSEPLEVVGAMVIRN 180

QY      181 SLTRGSAVFLVLEALTNFLNHRITPIVPLRGSISASGDLSPSLYTAAGITGPDVXX 240
DB      181 SLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISASGDLXKXLYAAILTGHPPXXKH 240

QY      241 VLESGEKKMPFRRBAISLFGAEAVVLGPKESLGLVNGTAVASAMATLSTLHDSHMLSLSQ 300
DB      241 XHEGEGEKIMKREBALXGLPEPVVGKESLGLVNGTAVASAMATLALHDAMLSLSQ 300

QY      301 ALTALTEAMVQGSFAPFIHDVCRPHPGQVEVARNIRTLTSSGSFAVEHEBEVXKXD 360
DB      301 ALTAETVEAMVGHAGSFXHFLHDVTRPHPTQIEVARIRITLGSXFAVHEBEVXKXD 360

QY      361 EGI LRDRPLRTSQPFI GFLVDEMMAHYSTLSJENNTDNDLLENKQKRAHGENPQA 420
DB      361 EGI LRDRPLRTSQMGPVSDMIFAHAAVLSLEASTDNDPLIVENKXTHGGENPQA 420

QY      421 SAYSIMETRIALALIKLNFOTCTELHNAAMNRGLPSCLAEDPSILNVHGKGLDHTA 480
DB      421 XAVANMETRIALALLIKGNPTQLETENLGNMNRGXSSCLAEDXSLSHGKGLDIAA 480

QY      481 AYASELGHANVYTFVQPAEMGQAVNSLALISARTAAANDVLSLLASHYCTLDAY 540
DB      481 AYTSELXHLANXYTHVOPREMGQAVNSLALISARXAXANDVLSLLATHYCLQAY 540

QY      541 DLAAMELDPKQFDLPTLLQOHLGTCLDVNALALEVKALNKRLEQTTVDLEPRMD 600
DB      541 DLAAMEFFKQPDPIXXXKXKXGHFGALDQXELXDKNMKXLEKRLQGNISYDLERMD 600

QY      601 AFSYATGVTELLSSSPSANTYLTAAVAMKVASAEKASISLTREVNRRFWQTPSSQAPAA 660
DB      601 AFSFATGVTELLSSSSXAKVSLAAVANAMKVASAEKASISLTRXVRDXFWAESSSPALX 660

QY      661 YLSPTRRVYSVRRELLGVQARRGDVFVQOETIGSVSRITYEALIKDGRINHVLKQMA 720
DB      661 YLSPTRRVYSVRRELLGVKARRGDVFVQOEVITIGVSRITYEALIKXIXINHVLKQMA 720

RESULT 6
US-09-939-408A-30
; Sequence 30, Application US/09939408A
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939, 408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624,693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 720

```

```
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Preferred
OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
NAME/KEY: misc feature
LOCATION: (5)..(15)
OTHER INFORMATION: The 'Xaa' at location 5 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (12)..(12)
OTHER INFORMATION: The 'Xaa' at location 12 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (16)..(16)
OTHER INFORMATION: The 'Xaa' at location 16 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (17)..(17)
OTHER INFORMATION: The 'Xaa' at location 17 stands for Thr, Ala, or Ser.
NAME/KEY: misc feature
LOCATION: (19)..(19)
OTHER INFORMATION: The 'Xaa' at location 19 stands for Gly.
NAME/KEY: misc feature
LOCATION: (20)..(20)
OTHER INFORMATION: The 'Xaa' at location 20 stands for a stop codon, Ser, or Leu.
NAME/KEY: misc feature
LOCATION: (25)..(25)
OTHER INFORMATION: The 'Xaa' at location 25 stands for Thr, Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (27)..(27)
OTHER INFORMATION: The 'Xaa' at location 27 stands for Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (28)..(28)
OTHER INFORMATION: The 'Xaa' at location 28 stands for Ala, or Pro.
NAME/KEY: misc feature
LOCATION: (34)..(34)
OTHER INFORMATION: The 'Xaa' at location 34 stands for Thr.
NAME/KEY: misc feature
LOCATION: (36)..(36)
OTHER INFORMATION: The 'Xaa' at location 36 stands for Arg, or Ser.
NAME/KEY: misc feature
LOCATION: (39)..(39)
OTHER INFORMATION: The 'Xaa' at location 39 stands for Ala, Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (40)..(40)
OTHER INFORMATION: The 'Xaa' at location 40 stands for Arg, Gly, or Trp.
NAME/KEY: misc feature
LOCATION: (48)..(48)
OTHER INFORMATION: The 'Xaa' at location 48 stands for Lys, Thr, Met, Glu, Ala, Val.
NAME/KEY: misc feature
LOCATION: (54)..(54)
OTHER INFORMATION: The 'Xaa' at location 54 stands for Val.
NAME/KEY: misc feature
LOCATION: (56)..(56)
OTHER INFORMATION: The 'Xaa' at location 56 stands for Lys, Glu, or Gln.
NAME/KEY: misc feature
LOCATION: (65)..(65)
OTHER INFORMATION: The 'Xaa' at location 65 stands for Glu, Asp, or Val.
NAME/KEY: misc feature
LOCATION: (66)..(66)
OTHER INFORMATION: The 'Xaa' at location 66 stands for Ile, Val, or Leu.
NAME/KEY: misc feature
LOCATION: (76)..(76)
OTHER INFORMATION: The 'Xaa' at location 76 stands for Gly.
NAME/KEY: misc feature
LOCATION: (87)..(87)
OTHER INFORMATION: The 'Xaa' at location 87 stands for Thr, Pro, or Ser.
NAME/KEY: misc feature
LOCATION: (93)..(93)
OTHER INFORMATION: The 'Xaa' at location 93 stands for Asp, or Ala.
NAME/KEY: misc feature
LOCATION: (102)..(102)
OTHER INFORMATION: The 'Xaa' at location 102 stands for Lys, or Asn.
NAME/KEY: misc feature
LOCATION: (103)..(103)
OTHER INFORMATION: The 'Xaa' at location 103 stands for Arg, or Ser.
NAME/KEY: misc feature
LOCATION: (109)..(109)
OTHER INFORMATION: The 'Xaa' at location 109 stands for Thr, Ala, or Ser.
NAME/KEY: misc feature
LOCATION: (112)..(112)
OTHER INFORMATION: The 'Xaa' at location 112 stands for Asp, His, or Tyr.
NAME/KEY: misc feature
LOCATION: (114)..(114)
OTHER INFORMATION: The 'Xaa' at location 114 stands for Arg, or Ser.
NAME/KEY: misc feature
LOCATION: (117)..(117)
OTHER INFORMATION: The 'Xaa' at location 117 stands for Gly.
NAME/KEY: misc feature
LOCATION: (148)..(148)
OTHER INFORMATION: The 'Xaa' at location 148 stands for Val.
NAME/KEY: misc feature
LOCATION: (150)..(150)
OTHER INFORMATION: The 'Xaa' at location 150 stands for Pro.
NAME/KEY: misc feature
LOCATION: (153)..(153)
OTHER INFORMATION: The 'Xaa' at location 153 stands for Ile, Val, or Phe.
NAME/KEY: misc feature
LOCATION: (154)..(154)
OTHER INFORMATION: The 'Xaa' at location 154 stands for Glu, or Asp.
NAME/KEY: misc feature
LOCATION: (157)..(157)
OTHER INFORMATION: The 'Xaa' at location 157 stands for Ser, Gly, or Arg.
NAME/KEY: misc feature
LOCATION: (159)..(159)
OTHER INFORMATION: The 'Xaa' at location 159 stands for Gly.
NAME/KEY: misc feature
LOCATION: (183)..(183)
OTHER INFORMATION: The 'Xaa' at location 183 stands for Thr.
NAME/KEY: misc feature
LOCATION: (223)..(223)
OTHER INFORMATION: The 'Xaa' at location 223 stands for Pro.
NAME/KEY: misc feature
LOCATION: (225)..(225)
OTHER INFORMATION: The 'Xaa' at location 225 stands for Ser.
NAME/KEY: misc feature
LOCATION: (237)..(237)
OTHER INFORMATION: The 'Xaa' at location 237 stands for Ser, Thr, Ile, Gly, Ala.
NAME/KEY: misc feature
LOCATION: (239)..(239)
OTHER INFORMATION: The 'Xaa' at location 239 stands for Val.
NAME/KEY: misc feature
LOCATION: (241)..(241)
OTHER INFORMATION: The 'Xaa' at location 241 stands for Val.
NAME/KEY: misc feature
LOCATION: (242)..(242)
OTHER INFORMATION: The 'Xaa' at location 242 stands for Val, Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (259)..(259)
OTHER INFORMATION: The 'Xaa' at location 259 stands for Leu, or Phe.
NAME/KEY: misc feature
LOCATION: (305)..(305)
OTHER INFORMATION: The 'Xaa' at location 305 stands for Leu.
NAME/KEY: misc feature
LOCATION: (319)..(319)
OTHER INFORMATION: The 'Xaa' at location 319 stands for Pro.
NAME/KEY: misc feature
LOCATION: (346)..(346)
OTHER INFORMATION: The 'Xaa' at location 346 stands for Lys, Arg, Thr, Gln, Pro, a
NAME/KEY: misc feature
LOCATION: (346)..(346)
OTHER INFORMATION: stop codon, Trp, or Ser.
```



```

/ LOCATION: (411)..(411)
/ OTHER INFORMATION: The 'Xaa' at location 411 stands for Lys, Thr, or Met.
/ NAME/KEY: misc feature
/ LOCATION: (421)..(421)
/ OTHER INFORMATION: The 'Xaa' at location 421 stands for Thr, Ala, or Ser.
/ NAME/KEY: misc feature
/ LOCATION: (457)..(457)
/ OTHER INFORMATION: The 'Xaa' at location 457 stands for Leu.
/ NAME/KEY: misc feature
/ LOCATION: (458)..(458)
/ OTHER INFORMATION: The 'Xaa' at location 458 stands for Pro.
/ NAME/KEY: misc feature
/ LOCATION: (466)..(466)
/ OTHER INFORMATION: The 'Xaa' at location 466 stands for Pro.
/ NAME/KEY: misc feature
/ LOCATION: (487)..(487)
/ OTHER INFORMATION: The 'Xaa' at location 487 stands for Gly.
/ NAME/KEY: misc feature
/ LOCATION: (493)..(493)
/ OTHER INFORMATION: The 'Xaa' at location 493 stands for Val.
/ NAME/KEY: misc feature
/ LOCATION: (518)..(518)
/ OTHER INFORMATION: The 'Xaa' at location 518 stands for Thr.
/ NAME/KEY: misc feature
/ LOCATION: (536)..(536)
/ OTHER INFORMATION: The 'Xaa' at location 536 stands for Val.
/ NAME/KEY: misc feature
/ LOCATION: (556)..(556)
/ OTHER INFORMATION: The 'Xaa' at location 556 stands for Met, Ile, Val, or Leu.
/ NAME/KEY: misc feature
/ LOCATION: (557)..(557)
/ OTHER INFORMATION: The 'Xaa' at location 557 stands for Ile, Val, or Leu.
/ NAME/KEY: misc feature
/ LOCATION: (558)..(558)
/ OTHER INFORMATION: The 'Xaa' at location 558 stands for Thr, Ala, or Pro.
/ NAME/KEY: misc feature

```

Query Match 75.6%; Score 2730; DB 24; Length 720;
 Best Local Similarity 76.5%; Pred. No. 3.7e-256;
 Matches 551; Conservative 45; Mismatches 124; Indels 0; Gaps 0;

```

1 MAPSLDLSLATTANGFTNGSHAAPTSAAGPTSAALRRPGLDGHAAHQSOUEIYOELISD 60
1 MAPSXDSINTSANGXNXXHAPKXXXTGATSLXLRPXLAAPATQXTQDIDEXILIAD 60
61 PDDVVELSGSYLTVRDVVGARKGRVRVQNDDEIRARVDKSVDFLKAQOLNSVGYTT 120
61 PDDDXELDGYTLTXDVVGARKGRVRVQTXDEIRAKIDXVEFLRXQLXNVVYVTT 120
121 GFGSADRTTEADAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMTVRVN 180
121 GFGSADRTTEADAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMTVRVN 180
181 SLTRGHSARVLYLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGATTGHPDVVH 240
181 SLTRGHSARVLYLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGATTGHPDVVH 240
241 VHEGTEKIMFARERALSIFGLEAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 VHEGTEKIMFARERALSIFGLEAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 XHEGTEKIMFARERALSIFGLEAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
301 ALTALTVEAMVQOQGSFAPFIHDCRPHFGQVEVARNRITLLSGSFVHEEVEVKYKD 360
301 ALTALTVEAMVQOQGSFAPFIHDCRPHFGQVEVARNRITLLSGSFVHEEVEVKYKD 360
361 EGLLRDRPLRTSPQFLGPIVEDMMAHAYSTLSLNTTNNPLIDVENKOTAHGNGFOA 420
361 EGLLRDRPLRTSPQFLGPIVEDMMAHAYSTLSLNTTNNPLIDVENKOTAHGNGFOA 420
421 SAVSISMEKTRIALALIGKLNFTQCTELNANMRGLSPCLAEDEPSLNYHGGLDIIH 480
421 SAVSISMEKTRIALALIGKLNFTQCTELNANMRGLSPCLAEDEPSLNYHGGLDIIH 480
421 XAVANTMEKTRIALALIGKLNFTQCTELNANMRGLSPCLAEDEPSLNYHGGLDIIH 480

```

```

QY 481 AYASELGHIANFVTTTFOVPAEMNGQAVNSLALISARPTAEANDVLSLLASHLYCTQAV 540
DB 481 AYASELGHIANFVTTTFOVPAEMNGQAVNSLALISARPTAEANDVLSLLASHLYCTQAV 540
QY 541 DTRAMELDKPKQPDPLPPLLQOHGTGIDVNAALAEYKALNKRLEOTTTDYDIEPRVHD 600
DB 541 DTRAMELDKPKQPDPLPPLLQOHGTGIDVNAALAEYKALNKRLEOTTTDYDIEPRVHD 600
QY 601 AFSVATGTVEILLSSSPSANVTLTZVNAMKVASAEKASITFREYANRPMQTPSSQAPAH 660
DB 601 AFSVATGTVEILLSSSPSANVTLTZVNAMKVASAEKASITFREYANRPMQTPSSQAPAH 660
QY 661 YLSPRTVLYSFVREELGVQARQGVFVQOETIGSNVSSIYEAIXKGRINHLVVKLA 720
DB 661 YLSPRTVLYSFVREELGVQARQGVFVQOETIGSNVSSIYEAIXKGRINHLVVKLA 720

```

RESULT 7
 PCT-US02-26934A-19
 Sequence 19, Application PC/TUS0226934A
 GENERAL INFORMATION:
 APPLICANT: PCB Services, Inc. et al.
 TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 TITLE OF INVENTION: Using Same
 FILE REFERENCE: 29479/50NSCA PCT
 CURRENT APPLICATION NUMBER: PCT/US02/26934A
 PRIOR FILING DATE: 2002-08-23
 PRIOR APPLICATION NUMBER: US 09/939,408
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 19
 LENGTH: 716
 TYPE: PRT
 ORGANISM: Rhodococcus toruloides
 PCT-US02-26934A-19

Query Match 72.4%; Score 2614; DB 1; Length 716;
 Best Local Similarity 72.0%; Pred. No. 9.5e-247;
 Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

1 MAPSLDLSLATTANGFTNGSHAAPTSAAGPTSAALRRPGLDGHAAHQSOUEIYOELISD 60
1 MAPSLDLSLHSHFANG-----VASAKQAVNGASNTNLAAGSHLPTTQVTVQDIEKMLAA 54
61 PDDVVELSGSYLTVRDVVGARKGRVRVQNDDEIRARVDKSVDFLKAQOLNSVGYTT 120
61 PDDDXELDGYTLTXDVVGARKGRVRVQTXDEIRAKIDXVEFLRXQLXNVVYVTT 120
121 GFGSADRTTEADAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMTVRVN 180
121 GFGSADRTTEADAVSLQKALIEHOLCGVTPTSXSPFSYGRGIENTLPLEVVRGAMTVRVN 180
181 SLTRGHSARVLYLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGATTGHPDVVH 240
181 SLTRGHSARVLYLEALTNFLNHRITPIVPLRGSISASGDSLPSYIAGATTGHPDVVH 240
241 VHEGTEKIMFARERALSIFGLEAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 VHEGTEKIMFARERALSIFGLEAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
241 XHEGTEKIMFARERALSIFGLEAVVLGPKEGLVNGTAVASAMATLSLHDSNMLSLISQ 300
301 ALTALTVEAMVQOQGSFAPFIHDCRPHFGQVEVARNRITLLSGSFVHEEVEVKYKD 360
301 ALTALTVEAMVQOQGSFAPFIHDCRPHFGQVEVARNRITLLSGSFVHEEVEVKYKD 360
361 EGLLRDRPLRTSPQFLGPIVEDMMAHAYSTLSLNTTNNPLIDVENKOTAHGNGFOA 420
361 EGLLRDRPLRTSPQFLGPIVEDMMAHAYSTLSLNTTNNPLIDVENKOTAHGNGFOA 420
421 SAVSISMEKTRIALALIGKLNFTQCTELNANMRGLSPCLAEDEPSLNYHGGLDIIH 479
421 SAVSISMEKTRIALALIGKLNFTQCTELNANMRGLSPCLAEDEPSLNYHGGLDIIH 479
421 XAVANTMEKTRIALALIGKLNFTQCTELNANMRGLSPCLAEDEPSLNYHGGLDIIH 479

```

```

QY      480 AAVASELGHLANPVTTFFVQPAEMNGQAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
DB      475 AAVTSELGHLANPVTTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLATHLYCYLOA 534
QY      540 VDLRAMELDFKKQDFPLPPLLQOHLG---TGLDV-NALALEYKALKKLEQTTYYLE 595
DB      535 IDLRALIEFEFKQFGPAIVSLIDQHFSGAMTGSNLRDELVEKVNKTLAKLEQTNNSYDLV 594
QY      596 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVYASAEKALSTREVRNRFMQPSSQ 655
DB      595 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVYASAEKALSTREVRNRFMQPSSQ 651
QY      656 APAAHAYLSPTRTVLVSFVREELGVQARRGDVFVQOQETIGSNVSRITYEAIKGRINHYL 715
DB      652 SPALSYLSPTRTQILVAFVREELGVQARRGDVFELGKQEVITIGSNVSRITYEAIKGRINNVL 711
QY      716 VKMLA 720
DB      712 LKMLA 716

RESULT 8
PCT-US03-05708-2
; Sequence 2, Application PC/TUS0305708
; GENERAL INFORMATION:
; APPLICANT: E.I. DuPont de Nemours and Company, Inc.
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: C11794 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/05708
; PRIOR FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
PCT-US03-05708-2

Query Match      72.4%; Score 2614; DB 1; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

```

QY      1 MAPSLDLSATTLANGFTNGSHAPTKSAGPTSAALRRTPGIDGHAHQSOLEIYQELISD 60
DB      1 MAPSLDISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQDIVEKMLAA 54
QY      61 PTDDVVELSGVSLTVRDVVGAAKGRVRVQNDDEIRAVDKSVDFPKAQLQNSVYGVTT 120
DB      55 PTSTSTLELDGYSNLGIVVSAARKGRPVAVKDSDEIRSKIDKSVFELRSQLSMSYGVTT 114
QY      121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSSGSFVGRGLENTLPLEVVRGAMVIRVN 180
DB      115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSFSPFLGRGLENSLPLEVVRGAMVIRVN 174
QY      181 SLTRGHSAYRLVVLALTNFLNHRITPIVPLRGISASGDSLPLYIAGAITGHPDVYK 240
DB      175 SLTRGHSAYRLVVLALTNFLNHRITPIVPLRGISASGDSLPLYIAGAITGHPDVYK 234
QY      241 VLHEGTEKIMFARBAISLFGLEAVVLGPKEGGLGVNGTAVASASMAATLSLHSHMLSLISQ 300
DB      235 VLHEGTEKILYAREEMALFNLEBPVVLGPKEGGLGVNGTAVASASMAATLSLHSHMLSLISQ 294
QY      301 ALTALTVEAMVGOQGSFAPFIHDVCRPHPGOVAVANRITLLSGSSFAVHEEVEKVKVD 360
DB      295 SLTAAITVEAMVGAHGSFHPFLADVTRPHPTQIEVAGNIRKLLBSGRFAVHEEVEKVKVD 354
QY      361 EGI LRQDRYPLRTSPQFGLPVEDMMAVSTLSLE-NNTTTDNPLLDVENKQTAAAGNFQ 419
DB      355 EGI LRQDRYPLRTSPQFGLPVEDMMAVSTLSLE-NNTTTDNPLLDVENKQTAAAGNFQ 414

```

```

QY      420 AAVSISMEKTRILALALGKLNFTQCTELTANANRGLPSCIAADEPSLNYHKGGLDIHT 479
DB      415 AAVANVMEKTRILGALQIGKLNFTQCTELTANANRGLPSCIAADEPSLNYHKGGLDIAA 474
QY      480 AAVASELGHLANPVTTFFVQPAEMNGQAVNSLALISARRTAENDVLSLLASHLYCTLOA 539
DB      475 AAVTSELGHLANPVTTTHVQPAEMANQAVNSLALISARRTTESNDVLSLLATHLYCYLOA 534
QY      540 VDLRAMELDFKKQDFPLPPLLQOHLG---TGLDV-NALALEYKALKKLEQTTYYLE 595
DB      535 IDLRALIEFEFKQFGPAIVSLIDQHFSGAMTGSNLRDELVEKVNKTLAKLEQTNNSYDLV 594
QY      596 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVYASAEKALSTREVRNRFMQPSSQ 655
DB      595 PRMHDAFSYATGTVVEILLSSPSANVTLTAVANAKVYASAEKALSTREVRNRFMQPSSQ 651
QY      656 APAAHAYLSPTRTVLVSFVREELGVQARRGDVFVQOQETIGSNVSRITYEAIKGRINHYL 715
DB      652 SPALSYLSPTRTQILVAFVREELGVQARRGDVFELGKQEVITIGSNVSRITYEAIKGRINNVL 711
QY      716 VKMLA 720
DB      712 LKMLA 716

```

```

RESULT 9
PCT-US03-17926-2
; Sequence 2, Application PC/TUS0317926
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: C1A912 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/17926
; PRIOR FILING DATE: 2003-05-20
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
PCT-US03-17926-2

Query Match      72.4%; Score 2614; DB 1; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

```

QY      1 MAPSLDLSATTLANGFTNGSHAPTKSAGPTSAALRRTPGIDGHAHQSOLEIYQELISD 60
DB      1 MAPSLDISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQDIVEKMLAA 54
QY      61 PTDDVVELSGVSLTVRDVVGAAKGRVRVQNDDEIRAVDKSVDFPKAQLQNSVYGVTT 120
DB      55 PTSTSTLELDGYSNLGIVVSAARKGRPVAVKDSDEIRSKIDKSVFELRSQLSMSYGVTT 114
QY      121 GFGGSADTRTEDAVSLQKALIEHQLCGVTPTXSSGSFVGRGLENTLPLEVVRGAMVIRVN 180
DB      115 GFGGSADTRTEDAISLQKALLEHQLCGVLPSFSPFLGRGLENSLPLEVVRGAMVIRVN 174
QY      181 SLTRGHSAYRLVVLALTNFLNHRITPIVPLRGISASGDSLPLYIAGAITGHPDVYK 240
DB      175 SLTRGHSAYRLVVLALTNFLNHRITPIVPLRGISASGDSLPLYIAGAITGHPDVYK 234
QY      241 VLHEGTEKIMFARBAISLFGLEAVVLGPKEGGLGVNGTAVASASMAATLSLHSHMLSLISQ 300
DB      235 VLHEGTEKILYAREEMALFNLEBPVVLGPKEGGLGVNGTAVASASMAATLSLHSHMLSLISQ 294
QY      301 ALTALTVEAMVGOQGSFAPFIHDVCRPHPGOVAVANRITLLSGSSFAVHEEVEKVKVD 360
DB      295 SLTAAITVEAMVGAHGSFHPFLADVTRPHPTQIEVAGNIRKLLBSGRFAVHEEVEKVKVD 354
QY      361 EGI LRQDRYPLRTSPQFGLPVEDMMAVSTLSLE-NNTTTDNPLLDVENKQTAAAGNFQ 419

```

```

Db      355  EQLIDQRYPLRTSPQWLGPIVSDLIHNAHVLTTAEQSTTDNPLDVENKTSHHGNGPQ 414
QY      420  ASAVSISNEKTRALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 479
Db      415  AAANVTMEKTRGLAQIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 474
QY      480  AAYASEIGHLANPVTTPQPAEMNGQAVNSLALISARTAEANDVLSLLASHTYCTLOA 539
Db      475  AAYTSELGHLANPVTTHQPAEMNGQAVNSLALISARTAESNDVLSLLATLHYCVLOA 534
QY      540  VDLRAMELDFKQDPPLPTLLQOHLG---TGLDV-NALALEYKALNKRLEQTTTYDLE 595
Db      535  IDLRAIEFEFKQGPALVSLIDQHFGSAMTGSNLRDELVEKNKTLAKRLEQTNISYDLY 594
QY      596  PRMHDAFSYATGTVELLSSSPSANVTTLTANAKVASAEKALSLTREVRNRPWQTPSSQ 655
Db      595  PRMHDAFSYATGTVEVLST---SLSLAANAKVAAASALSLTRQVETWMSAAS 651
QY      656  APAHAYLSPTRTVLYSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVL 715
Db      652  SPALSYLSPTRTQILYAFVREELGVKARRGDVFLGKQEVITIGSNVSKITYEAIKGRINVL 711
QY      716  VKMLA 720
Db      712  LKMLA 716

RESULT 10
US-09-939-408a-19
/ Sequence 19, Application US/09939408A
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ APPLICANT: Koester, Anna
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ TITLE OF INVENTION: Using Same
/ FILE REFERENCE: 29479/500NSCA
/ CURRENT APPLICATION NUMBER: US/09/939, 408A
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 09/624, 693
/ PRIOR FILING DATE: 2000-07-24
/ PRIOR APPLICATION NUMBER: PCT/US01/23270
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 19
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Rhodotorula toruloides
JS-09-939-408a-19

```

```

Query Match      72.4%, Score 2614; DB 24; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1  MAPSLDISLATTLANGFTNGSHAAPTSAAGPTSAALRPTGLDGHAAHQSGLTIVGELLSD 60
Db      1  MAPSLDISLISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVDIVERKMLAA 54
QY      61  PTDVVELSGSLTVPRDVVGAARKGRVRYONDEIRARVDSVDFKXQLONSYVGVT 120
Db      55  PTDSTLELDGSLNLDGVVSAARKGRPVRYKDSDELRSKIDKSVEFLRSQLSNYSYGVTT 114
QY      121  GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVSGLENTLPLEVVGAMTIRVN 180
Db      115  GFGGSADTRTEDAISTQKALIEHQLCGVLPSSPDSFRLRGLENSLPLEVVGAMTIRVN 174
QY      181  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGSISGDSLPLSTIAGAITGHPVKVN 240
Db      175  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGTISAGDSLPLSTIAGAITGHPDKVN 234
QY      241  VLHETEKIMPARBASLFGLEAVVLGPKKEGLGVNQTAVASAMATLSLHSHMLLSLQ 300

```

```

Db      225  VHEBKEKILYARBAKMLFLNIEPVVILGPKKEGLGVNQTAVASAMATLSLHSHMLLSLQ 294
QY      301  ALTALTYEAMVQOQSPAPFIHVCREPHQOVEARINRITLSGSSFAVHEBEVVKOD 360
Db      225  SLTAMTYEAMVGHAGSPHFLHDVTRPHPTQIEVAGNIRKLKSGSPFAVHEBEVVKOD 354
QY      361  EGLIRQDRYPLRTSPQWLGPIVSDLIHNAHVLTTAEQSTTDNPLDVENKTSHHGNGPQ 419
Db      355  EGLIRQDRYPLRTSPQWLGPIVSDLIHNAHVLTTAEQSTTDNPLDVENKTSHHGNGPQ 414
QY      420  ASAVSISNEKTRALALIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 479
Db      415  AAANVTMEKTRGLAQIGKLNFTQCTELNAAANRGLPSCLAEDPSLNYHKGDIH 474
QY      480  AAYASEIGHLANPVTTPQPAEMNGQAVNSLALISARTAEANDVLSLLASHTYCTLOA 539
Db      475  AAYTSELGHLANPVTTHQPAEMNGQAVNSLALISARTAESNDVLSLLATLHYCVLOA 534
QY      540  VDLRAMELDFKQDPPLPTLLQOHLG---TGLDV-NALALEYKALNKRLEQTTTYDLE 595
Db      535  IDLRAIEFEFKQGPALVSLIDQHFGSAMTGSNLRDELVEKNKTLAKRLEQTNISYDLY 594
QY      596  PRMHDAFSYATGTVELLSSSPSANVTTLTANAKVASAEKALSLTREVRNRPWQTPSSQ 655
Db      595  PRMHDAFSYATGTVEVLST---SLSLAANAKVAAASALSLTRQVETWMSAAS 651
QY      656  APAHAYLSPTRTVLYSFVREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIHVL 715
Db      652  SPALSYLSPTRTQILYAFVREELGVKARRGDVFLGKQEVITIGSNVSKITYEAIKGRINVL 711
QY      716  VKMLA 720
Db      712  LKMLA 716

RESULT 11
US-10-138-970A-4
/ Sequence 4, Application US/10138970A
/ GENERAL INFORMATION:
/ APPLICANT: Sarisiani, Sima
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei
/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ TITLE OF INVENTION: hydroxycinnamic Acid
/ FILE REFERENCE: CL1777
/ CURRENT APPLICATION NUMBER: US/10/138, 970A
/ PRIOR FILING DATE: 2002-07-23
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 4
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: C. violaceum
US-10-138-970A-4

```

```

Query Match      72.4%, Score 2614; DB 27; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

QY      1  MAPSLDISLATTLANGFTNGSHAAPTSAAGPTSAALRPTGLDGHAAHQSGLTIVGELLSD 60
Db      1  MAPSLDISLISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVDIVERKMLAA 54
QY      61  PTDVVELSGSLTVPRDVVGAARKGRVRYONDEIRARVDSVDFKXQLONSYVGVT 120
Db      55  PTDSTLELDGSLNLDGVVSAARKGRPVRYKDSDELRSKIDKSVEFLRSQLSNYSYGVTT 114
QY      121  GFGGSADTRTEDAVSLQKALIEHQLCGVTPTSXSSFSVSGLENTLPLEVVGAMTIRVN 180
Db      115  GFGGSADTRTEDAISTQKALIEHQLCGVLPSSPDSFRLRGLENSLPLEVVGAMTIRVN 174
QY      181  SLTGHSAVRLVLEALTNFLNHRITPIVPLRGSISGDSLPLSTIAGAITGHPVKVN 240

```

```

Db 175 SLTRGSAVRVLVLEALTNFLNNGITPIVLRGTTISASGDLSPISYIAAIIISGHPDSKVH 234
QY 241 VHEGTEKIMFARPAISLFGLEAVVLGPKREGGLVNGTAVASAMATLSLHDSHMLLSLQ 300
Db 235 VHEGTEKILYARERAMAFLNLEPVYLGPKREGGLVNGTAVASAMATLSLHDSHMLLSLQ 294
QY 301 ALTALTYEAMVGOOGSPAPFIHDVCRPHQGVAVANIRKTLSSGSAFAHEEYKXKD 360
Db 295 SLTMTYEAMVGAHGSFHPFLHDVTRPHPTQIEVAQNIKRLKLGSRPAVHHEEYKXKD 354
QY 361 EGI LRDRYPLRTSPQWGLPVLVDMMHAYSLSLLE-NNTTTDNPLLDVENKQTAAHGNFQ 419
Db 355 EGI LRDRYPLRTSPQWGLPVLVDMMHAYSLSLLE-NNTTTDNPLLDVENKQTAAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 479
Db 415 AAAYANMTEKTRIGLQIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 474
QY 480 AAAYASLGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAAYSELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534
QY 540 VDLRAVELDKKQDPDLLPTLLOHGLG---TGLDV-NALALEYKALNKRLEQTTTYDLE 595
Db 535 IDLRALIEFEFKQGPALIVSLIDQHFGSAMGNSLRDELVEKVKTKLAKLEQTNSTYDLY 594
QY 596 PRMHDAFSYATGTVEVLELSSSPSANVTLLTAVNAKVASAEKASISLREVRNRFMOTPSQ 655
Db 595 PRMHDAFSYATGTVEVLELSSSPSANVTLLTAVNAKVASAEKASISLREVRNRFMOTPSQ 651
QY 656 APAAHYLSPTRTVLISFVREELGVQARRGDVFGVQOETIGSNVSIYEAIDGRIINVL 715
Db 652 SPALSYLSPTRTQILVAFVREELGVQARRGDVFLGKQEVITIGSNVSKIYEAIKSGRIINVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

```

```

RESULT 12
US-10-188-523B-8
; Sequence 8, Application US/10188523B
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-188-523B-8

```

```

Query Match 72.4%; Score 2614; DB 27; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
QY 1 MAPSDLSLATTLANGFTNGSHAPTSAGPTSAARPTGDLGHAHQSGLIYQELLSLD 60
Db 1 MAPSDLSLISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54
QY 61 PRDDVVELISGVSLTRDVVGAARKGRVVRVONDEIDARVDKSDPLKALQKSVYGVTT 120
Db 55 PFDSTLELDGYSLNLGVDVVASARKGRPVVRVADSPDEIRSKIDKSVFLRSQLSVSVGVTT 114

```

```

QY 121 GFGGSADRTEDAYASLQKALIEHOLGVPTXSFSVGRGLNTPLLEVAGAMVIRVN 180
Db 115 GFGGSADRTEDAYASLQKALIEHOLGVPTXSFSVGRGLNTPLLEVAGAMVIRVN 174
QY 181 SLTRGSAVRVLVLEALTNFLNNGITPIVLRGTTISASGDLSPISYIAAIIISGHPDSKVH 240
Db 175 SLTRGSAVRVLVLEALTNFLNNGITPIVLRGTTISASGDLSPISYIAAIIISGHPDSKVH 234
QY 241 VHEGTEKIMFARPAISLFGLEAVVLGPKREGGLVNGTAVASAMATLSLHDSHMLLSLQ 300
Db 235 VHEGTEKILYARERAMAFLNLEPVYLGPKREGGLVNGTAVASAMATLSLHDSHMLLSLQ 294
QY 301 ALTALTYEAMVGOOGSPAPFIHDVCRPHQGVAVANIRKTLSSGSAFAHEEYKXKD 360
Db 295 SLTMTYEAMVGAHGSFHPFLHDVTRPHPTQIEVAQNIKRLKLGSRPAVHHEEYKXKD 354
QY 361 EGI LRDRYPLRTSPQWGLPVLVDMMHAYSLSLLE-NNTTTDNPLLDVENKQTAAHGNFQ 419
Db 355 EGI LRDRYPLRTSPQWGLPVLVDMMHAYSLSLLE-NNTTTDNPLLDVENKQTAAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 479
Db 415 AAAYANMTEKTRIGLQIGKLNFTQCTELLNAAAMRGLPSCLAEDPSLHYKGLDIHI 474
QY 480 AAAYASLGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAAYSELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534
QY 540 VDLRAVELDKKQDPDLLPTLLOHGLG---TGLDV-NALALEYKALNKRLEQTTTYDLE 595
Db 535 IDLRALIEFEFKQGPALIVSLIDQHFGSAMGNSLRDELVEKVKTKLAKLEQTNSTYDLY 594
QY 596 PRMHDAFSYATGTVEVLELSSSPSANVTLLTAVNAKVASAEKASISLREVRNRFMOTPSQ 655
Db 595 PRMHDAFSYATGTVEVLELSSSPSANVTLLTAVNAKVASAEKASISLREVRNRFMOTPSQ 651
QY 656 APAAHYLSPTRTVLISFVREELGVQARRGDVFGVQOETIGSNVSIYEAIDGRIINVL 715
Db 652 SPALSYLSPTRTQILVAFVREELGVQARRGDVFLGKQEVITIGSNVSKIYEAIKSGRIINVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

```

```

RESULT 13
US-10-188-523C-8
; Sequence 8, Application US/10188523C
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009-CIP
; CURRENT APPLICATION NUMBER: US/10/188,523C
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: MICROSOFT OFFICE 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-188-523C-8

```

```

Query Match 72.4%; Score 2614; DB 27; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;
QY 1 MAPSDLSLATTLANGFTNGSHAPTSAGPTSAARPTGDLGHAHQSGLIYQELLSLD 60
Db 1 MAPSDLSLISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVQVIVKMLAA 54

```

6 PTDVVAVLSYSTVTEVDVVGAAKRRKRVAVONDDDEIRARVVKSVDFPKAOLQNSVVGTT 120
 55 PTOSTLELDGYSINLSDDVVSAARKRPRVAVKSDSEIRKIKSVFEFLSQNSVYCVTT 114
 121 GFGGSADPTEDTDAVSIQKALIEHQLCGTPTSXSSFSVGRGLIENTLPLEVRGAMVIRN 180
 115 GFGGSADPTEDTDAISIQKALLEHQLCGVLPSSFDFSRIGRLGLENSLPLEVRGAMTIRVN 174
 181 SLRGRSAVAVLYLVALETNFLNHRITPIYPLRGSIASAGDLSPLSYAGAITGHPDYKH 240
 175 SLRGRSAVAVLYLVALETNFLNHCITPIYPLRGITISAGDLSPLSYAAIISGHPDKKH 234
 241 VLHETEKIMFAEEALSLEGLAVYLGPXEGIGLVNCTAVASAMATLSLHDSHTLSL9G 300
 235 VVHEGKEKILYAEENALFNLEPVLGPXEGIGLVNCTAVASAMATLADHAYLSL9G 294
 301 ALVALVEAMVQGGSGFAPFHIDVCRPHPGQVEVARNRTILSSGSFVAVEEYVKYKD 360
 295 SLTPALVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLBGSRAVVEEYVKYKD 354
 361 EGIIRDORYPLRTSPQFLPVEEDMMAVSTLSE--NNTTNDNPLDVENKQTAHGNGFQ 419
 355 EGIIRDORYPLRTSPQGLPVSOLDIHAHAVLTIAGSTINDNPLIDVENKTHSHGNGFQ 414
 420 ASAVSISMEKTRIALALISKLNFTQCTELLNAANRGLPSCLAEDPSLNVHGKLDIH 479
 415 AAAVANMTEKTRIGLAIQIKLNFOTLTMLNAGMKRGIPSCLAEDPSLSYHCKGLDIAA 474
 480 AAASSELGHANPVTFVQPAAMGQANSLATISARTAEANVLSLLASHLYCTQA 539
 475 AATSELGHANPVTTTHVQPAAMANOANSLATISARTTESNVLSLLATHLYCVLQA 534
 540 VDIRAMELDFKKQFDELLPTLLIQHLG--TGLDV-NALALEVKALNKRLBQTTTYDLE 595
 535 IDIRAIEFEFFKQFGPAIVSLIDQIFGSGAMGSMNRDELVEKVKTKLAEQNSYDLY 594
 596 PRHADAFSYATGTVVETLSSSPSAAVTLTAVNAMKVASAEALSLTEVNNRWQSPSSQ 655
 595 PRHADAFSYATGTVVEVLSST--SLSLAAVNAVMAAEASISLTQVVEETWSAASIS 651
 656 APAHAALSPRTVYSFVVEELGVQARRGDVFPVQOETTSNYSRIYEAIXDGRINHYL 715
 652 SPALSTLSRTQILVAFVVEELGVQARRGDVFLQGQVETIGSNYSKIYEAIXSGRLINHYL 711
 716 VKKLA 720
 712 LKILA 716
 RESULT 14
 US-10-374-366-2
 / Sequence 2, Application US/10374366
 / GENERAL INFORMATION:
 / APPLICANT: Tang, Xiao-Song
 / APPLICANT: Milano, Joseph
 / TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
 / FILE REFERENCE: CL11794 US NA
 / CURRENT APPLICATION NUMBER: US/10/374,366
 / CURRENT FILING DATE: 2003-02-26
 / PRIOR APPLICATION NUMBER: 60/360,279
 / PRIORITY FILING DATE: 2002-02-26
 / NUMBER OF SEQ ID NOS: 203
 / SOFTWARE: PatentIn version 3.2
 / SEQ ID NO 2
 / LENGTH: 716
 / TYPE: PRT
 / ORGANISM: Rhodospiridium glutinis
 / IS-10-374-366-2

Query Match	72.4%;	Score 2614;	DB 29;	Length 716;
Best Local Similarity	72.0%;	Pred. No. 9.5e-247;		
Matches 522;	Conservative 86;	Mismatches 103;	Indels 14;	Gaps 5;

```

QY 1 MABSDLSATTLANGFTNSHAPPKYSAAGPFSALRRPFGDGHAAHSCQEIYQELLSD 60
Db 1 MABSDLSISSIFANG-----VASAQOANGASTNLAVAGSHLPTTQYQYDIYERKMLAA 54
QY 61 PTDDVVELSGSYLTVDVYGAARKGRVYVQDDEIRARVDKSYVDFLYQKQNSVGYTT 120
Db 55 PLSSTLELDQSYLNLGDVYSAARKKRPVRVKQSDIEIRKIDKXVEFLRSQLSMNVGYTT 112
QY 121 GFGGSADTRTEDAVSLQKALLIEHOLCGVTPTXSXSFVSGGJENTLPLEYVRGAMVTRVN 180
Db 115 GFGGSADTRTEDASISQKALLIEHOLCGVLPSSFDFSGFRGGRGENSLPLEYVRGAMTRVN 174
QY 181 SLRGHSAVALVYLAELNTPLNHRITPEIYPRGSI SASGDSLPSLSYAGATGTHPDVKH 240
Db 175 SLRGHSAVALVYLAELNTPLNHRITPEIYELGTSISAGDLSPLSYAAASIGHPDVKH 233
QY 241 VLHEGTEKIMFAREASILFLEBAVYLGPKEGJGLVNGTAVASAMATSLHDSHMLSLSQ 300
Db 235 VHEGKEXKILYAREAMALFNLEPVYLGREGJGLVNGTAVASAMATLALHDAHMLSLSQ 299
QY 301 ALTALVEAMVGOQSSFAPIHDVCRPHPGQVEVARNITLISGSSFAVEHEEVEKYKD 360
Db 295 SLTAMTVEAMVHAGSFHFPLHDVTRPHPTQIEVAGNIRKLLGSRFAVHEEVEKYKD 354
QY 361 EGLRODRYVLRTPQFGLPYEDMKHASTISLE-NNTTNDNPLVYENQZAHGNGFO 415
Db 355 EGLRODRYPLRTPSPQWLPGLVSDLIHAAVLTLEAGOSTTNDNPLVYENQZSHHGNGFO 414
QY 420 ASAVSISMEKTRIALALLIGLNFTOCTEELNANMRGLPSCLAEDPELYNHGKGLDHI 479
Db 415 AAAYANIMTEKTRIGLAQIOGKLNFTQUTEMLANGMKRGPLSCLAEDPELSHCHGGLDIA 478
QY 480 AAAYASBELGIANPVTFVQPAEMGQAVNSLALISARRTAENDVLSILLASHLYCTLQA 533
Db 475 AAYTSELGHLANPVTHVQPAEMANQAVNSLALISARRTTESNDVLSILLATHLYCYLQA 534
QY 540 VDIRAMELDPKQFQDLPTLLQCHL---TGLDV-NALALEYKALKNLKEQTTDYDLE 599
Db 535 IDIRALEFEKKQFQGAIVSLIDQHFGSAMGTSNIRDELVERKNTLAKRIEQNNSYDLV 594
QY 596 PRMHDAFSYATGTVVELSSPSNAVTTTAVANAMVAAAEKASILTREVRRFPQTPSSQ 655
Db 595 PRMHDAFSFAGVIVEVLST---SLSLAANVAMVAAAEKASILTREVRRTPFSASTS 654
QY 656 APAHAAYISPTRVLYSFVFEELGVQARGDVPVVGQETTSNVSRIYEAIKGRINHVIL 715
Db 652 SPALSYLSPTQLLYFAVEELGVARGDVFELGQEWETTSINSKSIYEAIKSGRINNVIL 714
QY 716 VKYLA 720
Db 712 LKXLA 716

```

```

RESULT 15
US-10-439-478-2
; Sequence 2, Application US/10439478
; GENERAL INFORMATION:
; APPLICANT: E. I. du Pont de Nemours and Co.
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sariaslani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Varelil, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: C11912
; CURRENT APPLICATION NUMBER: US/10/439,478
; CURRENT FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2

```

LENGTH: 716
TYPE: PRT
ORGANISM: Rhodotorula glutinis
US-10-439-478-2

Query Match 72.4%; Score 2614; DB 30; Length 716;
Best Local Similarity 72.0%; Pred. No. 9.5e-247;
Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```
QY 1 MAPSLDSLATTANGFTNGSHAPPKSAAGPTSLAKRTPGIDGHAHQSLIEIVQELSD 60  
DB 1 MAPSLDSLASHFANG-----VASAKQVNGASTNLVAQSHLPTTQTVQDIVERKMLAA 54  
QY 61 PDDVVELSGVSLTVRDVYGAARKGRVAVONDEIRARVYKSVDFLKAQLQNSVYGVTT 120  
DB 55 PFDSTLELDGYSINIGDVYSAARKGRPRVYKSDSEIRSKIDKSVFRLRSQLSMSVYGVTT 114  
QY 121 GFGGSADTRTEDAVSLQALIEHQICGVTPTSXSSFSVGRGLENTLEPVVQAMVIRVN 180  
DB 115 GFGGSADTRTEDAISLQKALLEHQICGVLPSSFSFRLGRGLENTLEPVVQAMTIRVN 174  
QY 181 SLTRGHSAYRLVYVLEALNPNLNRTPVPLRGSIASGDLSPSYIAGAITHGHPDYKH 240  
DB 175 SLTRGHSAYRLVYVLEALNPNLNRTPVPLRGSIASGDLSPSYIAGIISGHPDSKYH 234  
QY 241 VLHEGTEKIMFARBAISLFGLEAVVLGPKBGLGVNGTAVASAMATLSLHDSHMLSLGQ 300  
DB 235 VYHEGTEKILYARBAALFNLEPVYLGPKBGLGVNGTAVASAMATLALHDAHMLSLGQ 294  
QY 301 ALTALTVEAMVQGGSSAFPHIDVCRPHPGQVEVARNIRTLSSSSFAVEHEEEVYKXD 360  
DB 295 SLTAMTVEAMVGHAGSFHFDVTRPHPTQIEVAGNIRKLLEGSRFAVHEEEVYKXD 354  
QY 361 EGI LRQDRYPLTSPOFGPLVEDMWHAYSTLSLE-NNTTDNPLDVENKQTAHGNGFQ 419  
DB 355 EGI LRQDRYPLTSPOFGPLVSDLIHNAVLTEAGOSTTNDPLIDVENKTSHGNGFQ 414  
QY 420 ASAAGISMEKTRIALALIGKLNFTQCTELMAANNRGLPSCLAAEDPSLNYHGKGLDIH 479  
DB 415 AAAVANMTWKTGLIAQIGKLNFTQCTELMAANNRGLPSCLAAEDPSLSYHGKGLDIA 474  
QY 480 AAYASLIGHLANPVTTFVOPAEQMGNOAVNSLALISARRTAENDVLSLLASHLYCTLOA 539  
DB 475 AAYTSELGHLANPVTTFVOPAEQMGNOAVNSLALISARRITESNDVLSLLATHLYCYLOA 534  
QY 540 VDLRAMELDKKQDFPLLPTLLOQHLG--TGLDV-NALALEVKKALNKRLEQTTTYDLE 595  
DB 535 IDLRAIEFEFFKQGFPAIVSLIDQFSGAMTGSNLRDELVEKVKNTLAKRLEQTNISDYLV 594  
QY 596 PRWHDAFSYATGTVELLSSSPSANVLTITANANKVASAKAISLTREVNRFWQTPSSQ 655  
DB 595 PRWHDAFSYATGTVEEVLST--SLSLAAVANANKVAAABSAISLTRQVRETFWSAASST 651  
QY 656 APAHAAYLSFRTVLVSFYREELGVQARRGDVFGVQOETIGSNVSRITYEAIKDGRIINHYL 715  
DB 652 SPALSYLSFRTOILYAFREELGVQARRGDVFLGQEVITIGSNVSKIYEAIKSGRIINHYL 711  
QY 716 VKMLA 720  
DB 712 LKMLA 716
```

Search completed: September 9, 2004, 10:19:37
Job time : 298.768 secs


```

Db      355 EGIIRORRPLRTSPQWLGPLVSDLIHHAVALTIEAGOSTTDNDLIVEKTBHGNFQ 414
QY      420 ASAASISMEKTRIALALIGKLNFTQCTELLNAAMNRLPSCLAEDPSLNHYGKLDIHI 479
Db      415 AAAYANWMEKTRIGLAQIGKLNFTQCTELMNAAMNRLPSCLAEDPSLNHYGKLDIAA 474
QY      480 AAAYASSELGHANPVTTPVQPAEMNGQAVNSALISARRTAEANDVLSILLASHYCTLOA 539
Db      475 AAYTSELGHANPVTTHVQPAEMANOAVNSALISARRTESNDVLSILLATHLYCVLOA 534
QY      540 VDLRAMELDPRKQDPDLLPTLLQOHLG---TGLDV-NALALEYKALNKLREQTTTYDLE 595
Db      535 IDLRALIEFERFKQGPALVSLIDQHFSGAMTGSNLRBELVEKVNKTAKRLEQNSYDLV 594
QY      596 PRMHDAFSYATGTVEELSSPSANVTITANNAKVASAEKASITREYRNRFWQTPSSQ 655
Db      595 PRMHDAFSYATGTVEELSSPSANVTITANNAKVASAEKASITREYRNRFWQTPSSQ 651
QY      656 APAAHYLSPTRTVLXSFYREELGVQARRGDVFGVQOETIGSNVSRITYEAIKGRINHYL 715
Db      652 SPALSYLSPTROIILYAFVREELGVARRGDVFLGKQEVITIGSNVSKIYEAIKGRINNYL 711
QY      716 VKMLA 720
Db      712 LKMLA 716

RESULT 2
US-10-476-198-4
/ Sequence 4, Application US/10476198
/ GENERAL INFORMATION:
/ APPLICANT: Saritsian, Sima
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei
/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ TITLE OF INVENTION: Hydroxycinnamic Acid
/ FILE REFERENCE: C11777
/ CURRENT APPLICATION NUMBER: US/10/476,198
/ CURRENT FILING DATE: 2003-10-24
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 4
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: C. violaceum
US-10-476-198-4

```

Query Match 72.4%; Score 2614; DB 6; Length 716;

Best Local Similarity 72.0%; Pred. No. 3.2e-194; Indels 14; Gaps 5;

Matches 522; Conservative 86; Mismatches 103; Indels 14; Gaps 5;

```

QY      1 MAPSDLSLATTLANGFTNGSHAAPTSAAGPTSAARTRTPGLDGHAAQSLIYQELSD 60
Db      1 MAPSDLSISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTQVTVQVDIVKMLAA 54
QY      61 PTDDVVELSGSYLTIRDVVGAARKGRARRVONDEIFARVDKSYDPLKALQONSVYGYTT 120
Db      55 PTDDVVELSGSYLTIRDVVGAARKGRARRVONDEIFARVDKSYDPLKALQONSVYGYTT 114
QY      121 GFGGSADTRTEDAVSLQKALIEHOLCGVTPTSXSSFSVGLENTLLPLEVVRAMVIRVN 180
Db      115 GFGGSADTRTEDAISLQKALIEHOLCGVLPSSPFSFLGLGELNSLPLEVVRAMVIRVN 174
QY      181 SLTRGSAVRLVLEALTNPLNKRITPIVPLKSGISASGDLSPSTIAGITGHPVAKH 240
Db      175 SLTRGSAVRLVLEALTNPLNKRITPIVPLKSGISASGDLSPSTIAGITGHPVAKH 234
QY      241 VHEGTEKIMFARBAISLFGLEAVVLGPKXELGLVNGTAVSASMATLSLSDSHMLSLIQ 300
Db      235 VHEGTEKILVARREMAALFNLEPVVLGPKXELGLVNGTAVSASMATLSLSDSHMLSLIQ 294
QY      301 ALTALTVEAMVGQGSFAFIHDVCRPHPGQVEVARNIRTLSSGSSFAVEHEEVKVKDD 360

```

```

Db      295 SLTMATVEANVAGHAGSPHPLHDVTRPHPTQGLEVAENIKRLLEGSPFAVHEEYKVD 354
QY      361 EGIIRORRPLRTSPQWLGPLVSDLIHHAVALTIEAGOSTTDNDLIVEKTBHGNFQ 419
Db      355 EGIIRORRPLRTSPQWLGPLVSDLIHHAVALTIEAGOSTTDNDLIVEKTBHGNFQ 414
QY      420 ASAASISMEKTRIALALIGKLNFTQCTELLNAAMNRLPSCLAEDPSLNHYGKLDIHI 479
Db      415 AAAYANWMEKTRIGLAQIGKLNFTQCTELMNAAMNRLPSCLAEDPSLNHYGKLDIAA 474
QY      480 AAAYASSELGHANPVTTPVQPAEMNGQAVNSALISARRTAEANDVLSILLASHYCTLOA 539
Db      475 AAYTSELGHANPVTTHVQPAEMANOAVNSALISARRTESNDVLSILLATHLYCVLOA 534
QY      540 VDLRAMELDPRKQDPDLLPTLLQOHLG---TGLDV-NALALEYKALNKLREQTTTYDLE 595
Db      535 IDLRALIEFERFKQGPALVSLIDQHFSGAMTGSNLRBELVEKVNKTAKRLEQNSYDLV 594
QY      596 PRMHDAFSYATGTVEELSSPSANVTITANNAKVASAEKASITREYRNRFWQTPSSQ 655
Db      595 PRMHDAFSYATGTVEELSSPSANVTITANNAKVASAEKASITREYRNRFWQTPSSQ 651
QY      656 APAAHYLSPTRTVLXSFYREELGVQARRGDVFGVQOETIGSNVSRITYEAIKGRINHYL 715
Db      652 SPALSYLSPTROIILYAFVREELGVARRGDVFLGKQEVITIGSNVSKIYEAIKGRINNYL 711
QY      716 VKMLA 720
Db      712 LKMLA 716

```

RESULT 3

US-10-476-193-19

/ Sequence 19, Application US/10476193

/ GENERAL INFORMATION:

/ APPLICANT: Saritsian, Sima

/ APPLICANT: Tang, Xiao-Song

/ APPLICANT: Qi, Wei Wei

/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

/ TITLE OF INVENTION: Hydroxycinnamic Acid

/ FILE REFERENCE: C11777

/ CURRENT APPLICATION NUMBER: US/10/476,193

/ CURRENT FILING DATE: 2003-10-28

/ NUMBER OF SEQ ID NOS: 24

/ SOFTWARE: Microsoft Office 97

/ SEQ ID NO 19

/ LENGTH: 716

/ TYPE: PRT

/ ORGANISM: Rhodotorula glutinis

/ FEATURE:

/ NAME/KEY: MISC FEATURE

/ LOCATION: (198)..(198)

/ OTHER INFORMATION: X=Asp, Asn, Glu, Gln

/ FEATURE:

/ NAME/KEY: MISC FEATURE

/ LOCATION: (198)..(198)

/ OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln

/ FEATURE:

/ NAME/KEY: MISC FEATURE

/ LOCATION: (540)..(540)

/ OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly

/ US-10-476-193-19

Query Match 72.4%; Score 2613; DB 6; Length 716;

Best Local Similarity 72.0%; Pred. No. 3.8e-194; Indels 14; Gaps 5;

Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

```

QY      1 MAPSDLSLATTLANGFTNGSHAAPTSAAGPTSAARTRTPGLDGHAAQSLIYQELSD 60
Db      1 MAPSDLSISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTQVTVQVDIVKMLAA 54
QY      61 PTDDVVELSGSYLTIRDVVGAARKGRARRVONDEIFARVDKSYDPLKALQONSVYGYTT 120

```

```

Db      55 PDSTSELGYSINLDGVSAARKGPVRVKDSDIRKIKSVFELSQLSMSYGVTT
      121 GFSGADRTTEDAVSLQKALIEHOLCGVPTXSFSFSGRGENTLPLEVRGAMVRVN
      115 GFSGADRTTEDAISLQKALLEHQLCGVLPSSFDSFRLGRLENSLPLEVRGAMTVRN
      181 SLTRGSAVRVLVLEALNPLNHRITPIVPLRGSIASGDSLPSLYAGAITGHPDVKH
      175 SLTRGSAVRVLVLEALNPLNHRITPIVPLRGTSASGDSLPSLYAALISGHPDSKVH
      241 VHEGTEKIMFARREALSLFGLBAVVLGPKREGGLVNGTAVASAMATLSLHSHLSLSQ
      235 VAEHGEKELIYAREALNLEPVLGPKREGGLVNGTAVASAMATLSLHSHLSLSQ
      301 ALTALIVEAMVQOQSFPAPFIHDVCRPHPGQVEVARNIRTLISGSSFAVEHEEVKYKD
      295 SLTAMTVEAMVGHASGFHPLHDVTRPHPTQIEVAGNIRKLLEGSRAVHHEEVKYKD
      361 EGLRQDRYPLRTSPQFLGPIVEDMMAHASTLSLE--NNTTNDNPLLDVENKQTAHGNFQ
      355 EGLRQDRYPLRTSPQFLGPIVEDMMAHASTLSLE--NNTTNDNPLLDVENKQTAHGNFQ
      420 ASAVSISWEKTRIALALIGKLNFTQCTELLNAAMNRGLPSCLAEDPRLNHYGGLDHI
      415 AAVANTMEKTRIGLAQIGKLNFTQCTELLNAAMNRGLPSCLAEDPRLNHYGGLDIAA
      480 AAVASLGHIANPVTTFVOPAEKNOAVNSLALISARTAEANDVLSLLASHLYCTLOA
      475 AAVTSELGHIANPVTTFVOPAEKNOAVNSLALISARTAEANDVLSLLASHLYCTLOA
      540 VDIRAMELDFKKQFDPPLPTLIQOHLG--TGLDV--NALALEVKKALKRLEQTTTDDLE
      535 IDIRAXEFEFFKQFGPAIVSLIDQHFGSAMTGSNIRDELVEKVKNTLAKRLEQNTSYDLV
      596 PRMHDAFSYATGTVEELSSSPSANVTLTAVAMKVASAEKALSLTRVVRNRFQTPSSQ
      595 PRMHDAFSYATGTVEELSSSPSANVTLTAVAMKVASAEKALSLTRVVRNRFQTPSSQ
      656 APHAHALSRTKVLVSFYAREELGVQARRGDVFGVQOETIGSNVSRILEAKDGRINHL
      652 SPALSTLSRTOILVAFVAREELGVQARRGDVFGVQOETIGSNVSRILEAKDGRINHL
      716 VKMLA 720
      712 LKMLA 716

Db      716 VKMLA 720
      712 LKMLA 716

RESULT 4
US-10-476-198-19
; Sequence 19, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Satiastani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: CL1777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (198)..(198)

```

```

; OTHER INFORMATION: Xaa=Asp, Asn, Glu, Gln
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-198-19

Query Match      72.4%; Score 2613; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 3,8e-194;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

      1 MPBLSLATTIANFTNGSHAPRTKSAAGPTSLARRPGLDGHAHQSOCEIYQELISD
      1 MAPSLDSISHSPANG-----VAAKQAVNGASTNLAAVAGSHLPTTQYTVDIYKMLAA
      61 PDDVVELSGEVLTVRDVYGAAKRGRRVONDEIRARVDKSVPLKAOLONSYGVTT
      55 PDSTSELGYSINLDGVSAARKGPVRVKDSDIRKIKSVFELSQLSMSYGVTT
      121 GFSGADRTTEDAVSLQKALIEHOLCGVPTXSFSFSGRGENTLPLEVRGAMVRVN
      115 GFSGADRTTEDAISLQKALLEHQLCGVLPSSFDSFRLGRLENSLPLEVRGAMTVRN
      181 SLTRGSAVRVLVLEALNPLNHRITPIVPLRGSIASGDSLPSLYAGAITGHPDVKH
      175 SLTRGSAVRVLVLEALNPLNHRITPIVPLRGTSASGDSLPSLYAALISGHPDSKVH
      241 VHEGTEKIMFARREALSLFGLBAVVLGPKREGGLVNGTAVASAMATLSLHSHLSLSQ
      235 VAEHGEKELIYAREALNLEPVLGPKREGGLVNGTAVASAMATLSLHSHLSLSQ
      301 ALTALIVEAMVQOQSFPAPFIHDVCRPHPGQVEVARNIRTLISGSSFAVEHEEVKYKD
      295 SLTAMTVEAMVGHASGFHPLHDVTRPHPTQIEVAGNIRKLLEGSRAVHHEEVKYKD
      361 EGLRQDRYPLRTSPQFLGPIVEDMMAHASTLSLE--NNTTNDNPLLDVENKQTAHGNFQ
      355 EGLRQDRYPLRTSPQFLGPIVEDMMAHASTLSLE--NNTTNDNPLLDVENKQTAHGNFQ
      420 ASAVSISWEKTRIALALIGKLNFTQCTELLNAAMNRGLPSCLAEDPRLNHYGGLDHI
      415 AAVANTMEKTRIGLAQIGKLNFTQCTELLNAAMNRGLPSCLAEDPRLNHYGGLDIAA
      480 AAVASLGHIANPVTTFVOPAEKNOAVNSLALISARTAEANDVLSLLASHLYCTLOA
      475 AAVTSELGHIANPVTTFVOPAEKNOAVNSLALISARTAEANDVLSLLASHLYCTLOA
      540 VDIRAMELDFKKQFDPPLPTLIQOHLG--TGLDV--NALALEVKKALKRLEQTTTDDLE
      535 IDIRAXEFEFFKQFGPAIVSLIDQHFGSAMTGSNIRDELVEKVKNTLAKRLEQNTSYDLV
      596 PRMHDAFSYATGTVEELSSSPSANVTLTAVAMKVASAEKALSLTRVVRNRFQTPSSQ
      595 PRMHDAFSYATGTVEELSSSPSANVTLTAVAMKVASAEKALSLTRVVRNRFQTPSSQ
      656 APHAHALSRTKVLVSFYAREELGVQARRGDVFGVQOETIGSNVSRILEAKDGRINHL
      652 SPALSTLSRTOILVAFVAREELGVQARRGDVFGVQOETIGSNVSRILEAKDGRINHL
      716 VKMLA 720
      712 LKMLA 716

Db      716 VKMLA 720
      712 LKMLA 716

RESULT 5
US-10-476-193-24
; Sequence 24, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Satiastani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

```

FILE REFERENCE: CL1777
CURRENT APPLICATION NUMBER: US/10/476,193
CURRENT FILING DATE: 2003-10-28
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 24
LENGTH: 716
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-476-193-24

Query Match 72.4%; Score 2612; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 4,66-194;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSATLTANGFTNGSHAPTKSAAGPTSAARPTGDLGHAHQSLQLEIYQELSD 60
DB 1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQYTVQVIVKXMLAA 54
QY 61 PTDDVVELSGYSLTVRDVYGAARKGRVRYVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTDSLTLEDGYSNLNDGVYSAARKGRPVRYVKDSDEIRSKIDKSVDFLSQLSMSVYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALLIEHQLCGVTPTSXSESVGRLGENTLPLEVVRGAMVTRVN 180
DB 115 GFGGSADRTEDAVSLQKALLIEHQLCGVTPSFSFRLGRGLGENTLPLEVVRGAMVTRVN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGITGHPDVYKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAAISGHPDSKVH 234
QY 241 VHEGTEKIMFARBAISLFGLEAVVLPKRGGLGVNGTAVASAMATLSLHDSHMLSLLSQ 300
DB 235 VHEGKEKILYAREEMALFNLEPVLPKRGGLGVNGTAVASAMATLSLHDAHMLSLLSQ 294
QY 301 ALTALTVEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
DB 295 SLTRMTVEAMVGAGSFHPLHDVTRPHPTQLEVAGNIRKLEGSRAVAHHEEVKXKD 354
QY 361 EGILRODRYPLRTSPQFIPLGPIVEDMKAAYSTLSLNTTNDPLLDVENKOTAHGQNFQ 419
DB 355 EGILRODRYPLRTSPQFIPLGPIVEDMKAAYSTLSLNTTNDPLLDVENKOTAHGQNFQ 414
QY 420 ASAVSISMERTKRLALIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIAA 474
DB 415 AAIVANTMERKRLGLAIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIAA 474
QY 480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHLANPVTTHQPAEMGNQAVNSLALISARPTESNDVLSLLATHLYCTLOA 534
QY 540 VDIRAMELDFKQDPDLPLTLLOQHIG--TGLDV-NALALEVKKALKRLEQTTTYDLE 595
DB 535 IDLRATEPEFKQDPQPAIVSLIDQHFSGAMTGSNRLDEIVEKNKTLAKRLEQNTSYDLV 594
QY 596 PRHMDAFSVATGTVVLELSSPSANVTLLAVANMKAASAKAISTREYANRFPQTPSSQ 655
DB 595 PRHMDAFSVATGTVVLELSSPSANVTLLAVANMKAASAKAISTREYANRFPQTPSSQ 651
QY 656 APAHAAYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHYL 715
DB 652 SPALSYLSPTRTQILYAFVREELGVKARRGDVFLGQEVITIGSNVSKIYEAIKGRINHYL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

RESULT 6
US-10-476-198-24
; Sequence 24, Application US/10476198

GENERAL INFORMATION:
APPLICANT: Sartasiani, Sima
APPLICANT: Tang, Xiao-Song
APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the production of Tyrosine, Cinnamic Acid and Para-
TITLE OF INVENTION: Hydroxycinnamic Acid
FILE REFERENCE: CL1777
CURRENT APPLICATION NUMBER: US/10/476,198
CURRENT FILING DATE: 2003-10-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 24
LENGTH: 716
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-476-198-24

Query Match 72.4%; Score 2612; DB 6; Length 716;
Best Local Similarity 72.0%; Pred. No. 4,66-194;
Matches 522; Conservative 85; Mismatches 104; Indels 14; Gaps 5;

QY 1 MAPSLDLSATLTANGFTNGSHAPTKSAAGPTSAARPTGDLGHAHQSLQLEIYQELSD 60
DB 1 MAPSLDLSHSHFANG-----VASAKQAVNGASTNLAVAGSHLPTTQYTVQVIVKXMLAA 54
QY 61 PTDDVVELSGYSLTVRDVYGAARKGRVRYVQNDDEIRAVDKSVDFLKAQLQNSVYGVTT 120
DB 55 PTDSLTLEDGYSNLNDGVYSAARKGRPVRYVKDSDEIRSKIDKSVDFLSQLSMSVYGVTT 114
QY 121 GFGGSADRTEDAVSLQKALLIEHQLCGVTPTSXSESVGRLGENTLPLEVVRGAMVTRVN 180
DB 115 GFGGSADRTEDAVSLQKALLIEHQLCGVTPSFSFRLGRGLGENTLPLEVVRGAMVTRVN 174
QY 181 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAGITGHPDVYKH 240
DB 175 SLTRGSAVRLVLEALTNFLNHRITPIVPLRGSISAGDLSPLSYIAAISGHPDSKVH 234
QY 241 VHEGTEKIMFARBAISLFGLEAVVLPKRGGLGVNGTAVASAMATLSLHDSHMLSLLSQ 300
DB 235 VHEGKEKILYAREEMALFNLEPVLPKRGGLGVNGTAVASAMATLSLHDAHMLSLLSQ 294
QY 301 ALTALTVEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEVKXKD 360
DB 295 SLTRMTVEAMVGAGSFHPLHDVTRPHPTQLEVAGNIRKLEGSRAVAHHEEVKXKD 354
QY 361 EGILRODRYPLRTSPQFIPLGPIVEDMKAAYSTLSLNTTNDPLLDVENKOTAHGQNFQ 419
DB 355 EGILRODRYPLRTSPQFIPLGPIVEDMKAAYSTLSLNTTNDPLLDVENKOTAHGQNFQ 414
QY 420 ASAVSISMERTKRLALIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIAA 474
DB 415 AAIVANTMERKRLGLAIGKLNFTQCTELNANMRGLPSCLAADPSLSYHCKGLDIAA 474
QY 480 AAYASELGHLANPVTTFVQPAEMGNQAVNSLALISARPTAEANDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHLANPVTTHQPAEMGNQAVNSLALISARPTESNDVLSLLATHLYCTLOA 534
QY 540 VDIRAMELDFKQDPDLPLTLLOQHIG--TGLDV-NALALEVKKALKRLEQTTTYDLE 595
DB 535 IDLRATEPEFKQDPQPAIVSLIDQHFSGAMTGSNRLDEIVEKNKTLAKRLEQNTSYDLV 594
QY 596 PRHMDAFSVATGTVVLELSSPSANVTLLAVANMKAASAKAISTREYANRFPQTPSSQ 655
DB 595 PRHMDAFSVATGTVVLELSSPSANVTLLAVANMKAASAKAISTREYANRFPQTPSSQ 651
QY 656 APAHAAYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKGRINHYL 715
DB 652 SPALSYLSPTRTQILYAFVREELGVKARRGDVFLGQEVITIGSNVSKIYEAIKGRINHYL 711
QY 716 VKMLA 720
DB 712 LKMLA 716

Db 712 LKMLA 716

RESULT 7
US-10-476-193-22
; Sequence 22, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sartasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE OF INVENTION: Hydroxycinnamic Acid
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,193
; CURRENT FILING DATE: 2003-10-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: x=Gly, Ala, Ser,Thr, Pro
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: x=Thr, Ala, Ser, Pro, Gly
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-193-22

Query Match 72.2%; Score 2607; DB 6; Length 716;
Best Local Similarity 71.9%; Pred.No.1,1e-193;
Matches 521; Conservative 85; Mismatches 105; Indels 14; Gaps 5;

Db 1 MAPSLDSLATTLANFTNGSHAAPTKSAAGPTSAARRTPGDLGHAHOSOLEIYQELSLD 60
1 MAPSLDSLISHSPANG-----VASAKQAVNGASTNLVAAGSHLPTTQVTOYDIVERKMLAA 54
QY 61 PDDVVEISGYSLTVRDVYGAARKGRVRVQNDDEIRARVDSVDFKAQONSRYGVT 120
Db 55 PFDSTLELDGYSINLGDVVSAAKRGPRVVKDSDEIRKIDKSVDFRLSOLSMGVYGVTT 114
QY 121 GFGSADRTETDAVSLQKALIEHQICGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db 115 GFGSADRTETDAVSLQKALIEHQICGVLPSSPDSFRLGRGLENSLPLEVVRGAMTIRVN 174
QY 181 SLTRGHSAAVRLVLEALTNFNLNHRITPIVPLRGSSISASGDSLPSLYIAGAITGHPDKVA 240
Db 175 SLTRGHSAAVRLVLEALTNFNLNHRITPIVPLRGSSISASGDSLPSLYIAAISHGPDKNH 234
QY 241 VHEGTEKIMPARREASISLFGLEAVVLAGPKEGLGVNCTAVASASMATLSLHSHMLSLISQ 300
Db 235 VHEGTEKILYAREAMALFNLEPVVLGPKEGLGVNCTAVASASMATLALHDAHMLSLISQ 294
QY 301 ALTALTYEAMVGOOGSFAPIFDVCAEPHGOVEVARNIRTLISGSSFAVEHEEYKAKD 360
Db 295 SLTAMTYEAMVGHAGSFHFLDVTRRPHPTQIEVAGNIRKILLEGSRFAVHEEYKAKD 354
QY 361 EGLIRODRYPLRTSPQIFGLPVEDNMMAVSTLSLE-NTTTDNPLLDVENKQTAHGANFQ 419
Db 355 EGLIRODRYPLRTSPQIFGLPVEDNMMAVSTLSLE-NTTTDNPLLDVENKSHHGNGFQ 414
QY 420 AASVASTSMETRLALILKLNFTOCTELNNAANRGIPSCIAADPSLANTHGGIDHI 479
Db 415 AAVVANITMETRIGLQIGKLNFTOTTELTNKNMRGLPSCIAADPSLANTHGGIDHI 474
QY 480 AAVASELGHIANPVTTTFVPAEMGNQAVNSLLAISARTAEANDVLSLLASHLYCTLOA 539

Db 475 AAVTSELGHIANPVTTTHVQAPAMNQAAXNSLALISARTTESNDVLSLLATHLYCTLOA 534

QY 540 VDLRAMELDFKQEPDPLPTLIQCHLG---TGLDV-NALALEYKALKRLEQTTTDL 595
Db 535 IDLRAXEPEEFKQFQPAVAVSLIDQHFSGAMTGSNLRBELVKNKTLAKRLEQTNSSDLV 594
QY 596 PRWDAFSYATGTVELLSSPSANVTLTAVANAKVSAEKALSLTEVRNRFMQTESQ 655
Db 595 PRWDAFSYATGTVELLSSPSANVTLTAVANAKVSAEKALSLTEVRNRFMQTESQ 651
QY 656 APAAVLSFRTVLYSFVREBELGVQARRGVFVQOCTISSNSRIYEALKGRINHV 715
Db 652 SPALYSFRTQILYAFVREBELGVQARRGVFVQOCTISSNSRIYEALKGRINNV 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 8
US-10-476-198-22
; Sequence 22, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sartasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,198
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: x=Gly, Ala, Ser,Thr, Pro
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: x=Thr, Ala, Ser, Pro, Gly
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-22

Query Match 72.2%; Score 2607; DB 6; Length 716;
Best Local Similarity 71.9%; Pred.No.1,1e-193;
Matches 521; Conservative 85; Mismatches 105; Indels 14; Gaps 5;

QY 1 MAPSLDSLATTLANFTNGSHAAPTKSAAGPTSAARRTPGDLGHAHOSOLEIYQELSLD 60
1 MAPSLDSLISHSPANG-----VASAKQAVNGASTNLVAAGSHLPTTQVTOYDIVERKMLAA 54
QY 61 PDDVVEISGYSLTVRDVYGAARKGRVRVQNDDEIRARVDSVDFKAQONSRYGVT 120
Db 55 PFDSTLELDGYSINLGDVVSAAKRGPRVVKDSDEIRKIDKSVDFRLSOLSMGVYGVTT 114
QY 121 GFGSADRTETDAVSLQKALIEHQICGVTPTSXSSFSVGRGLENTLPLEVVRGAMVIRVN 180
Db 115 GFGSADRTETDAVSLQKALIEHQICGVLPSSPDSFRLGRGLENSLPLEVVRGAMTIRVN 174
QY 181 SLTRGHSAAVRLVLEALTNFNLNHRITPIVPLRGSSISASGDSLPSLYIAGAITGHPDKVA 240
Db 175 SLTRGHSAAVRLVLEALTNFNLNHRITPIVPLRGSSISASGDSLPSLYIAAISHGPDKNH 234
QY 241 VHEGTEKIMPARREASISLFGLEAVVLAGPKEGLGVNCTAVASASMATLSLHSHMLSLISQ 300
Db 235 VHEGTEKILYAREAMALFNLEPVVLGPKEGLGVNCTAVASASMATLALHDAHMLSLISQ 294
QY 301 ALTALTYEAMVGOOGSFAPIFDVCAEPHGOVEVARNIRTLISGSSFAVEHEEYKAKD 360
Db 295 SLTAMTYEAMVGHAGSFHFLDVTRRPHPTQIEVAGNIRKILLEGSRFAVHEEYKAKD 354
QY 361 EGLIRODRYPLRTSPQIFGLPVEDNMMAVSTLSLE-NTTTDNPLLDVENKQTAHGANFQ 419
Db 355 EGLIRODRYPLRTSPQIFGLPVEDNMMAVSTLSLE-NTTTDNPLLDVENKSHHGNGFQ 414
QY 420 AASVASTSMETRLALILKLNFTOCTELNNAANRGIPSCIAADPSLANTHGGIDHI 479
Db 415 AAVVANITMETRIGLQIGKLNFTOTTELTNKNMRGLPSCIAADPSLANTHGGIDHI 474
QY 480 AAVASELGHIANPVTTTFVPAEMGNQAVNSLLAISARTAEANDVLSLLASHLYCTLOA 539

Db 235 VHEGKEKILYAREBAMLFLEPVVLGPKKGLGVNGTAVSASMTALHDAHMLSLSQ 294

Qy 301 ALTALTEAMVGOOGSPAFPIHDVCRPHPGQVEVARINIRTLSSGSPAVEHEEVKXKD 360

Db 295 SLTAMTEAMVGHAGSFHPIHDVTRPHPTQIEVAGNIRKLEGSRFVAHHEEVKXKD 354

Qy 361 EGIIRQDRYPLRTSPQIGPLVEDMMAVSTLSLE-NNTTNDNPLLDVENKOTAHGNGFQ 419

Db 355 EGIIRQDRYPLRTSPQIGPLVEDMMAVSTLSLE-NNTTNDNPLLDVENKOTAHGNGFQ 414

Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNMAANRGLPSCLAEDPSLNYHGKGLDIHI 479

Db 415 AAAYANTMEKTRIGLAQIGKLNFTQCTELNMAANRGLPSCLAEDPSLNYHGKGLDIHI 474

Qy 480 AAYSELGHLANPVTTTPVQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539

Db 475 AAYSELGHLANPVTTTPVQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534

Qy 540 VDLRAEMLDFKQDPFLPTLLQCHLG---TGLDV-NALALEVKAANKRLBQTTDYLE 595

Db 535 IDLRAXEPEFKQGFPAIVSLIDQFGSAMTGSNLDELVEKVKTKLAKRLBQTNISDYLV 594

Qy 596 PRMHDAFSYATGTVELLSSSPSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSSQ 655

Db 595 PRMHDAFSYATGTVELLSSSPSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSSQ 651

Qy 656 APAAHYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKDGRIHVL 715

Db 652 SPALSYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKDGRIHVL 711

Qy 716 VKMLA 720

Db 712 LKMLA 716

RESULT 9

US-10-476-193-20

Sequence 20, Application US/10476193

GENERAL INFORMATION:

APPLICANT: Saitasian, Sima

APPLICANT: Tang, Xiao-Song

APPLICANT: Qi, Wei Wei

TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

FILE REFERENCE: C11777

CURRENT APPLICATION NUMBER: US/10/476,193

CURRENT FILING DATE: 2003-10-28

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

SEQ ID NO 20

LENGTH: 716

TYPE: PRT

ORGANISM: Rhodotorula glutinis

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (181)..(181)

OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (235)..(235)

OTHER INFORMATION: X=Ala, Gly, Ser, Thr, Pro

FEATURE:

NAME/KEY: MISC_FEATURE

LOCATION: (540)..(540)

OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly

NAME/KEY: MISC_FEATURE

LOCATION: (540)..(540)

OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly

US-10-476-193-20

Query Match 72.1%; Score 2603; DB 6; Length 716;

Best Local Similarity 71.7%; Pred. No. 2.3e-193;

Matches 520; Conservative 85; Mismatches .106; Indels 14; Gaps 5;

Qy 1 MAPSIDSLATTLANGFTNGSHAAPTYSAGPTSAIRRTPGIDGHAHQOLEIVQELISD 60

Db 1 MAPSIDSLISHSPANG-----VASAKQAVNGASTNLVAGSHLPTTQVTDIVKMLAA 54

Qy 61 PDDVVEISGVSLTVDVVGAAKGRARYONDEIPARVYKSDVFKACLOQNSVYGVTT 120

Db 55 PTDSTIELDGLYSLNIGDVSAKGRPPYKRVKDSIISKIDKSEFIRLSQNSVYGVTT 114

Qy 121 GGGASDRTTEDAVSLQALIEHQLCGYTPTSXSSFSVGRGLENTLEEVVGMVIRVN 180

Db 115 GGGASDRTTEDAVSLQALIEHQLCGYTPTSXSSFSVGRGLENTLEEVVGMVIRVN 174

Qy 181 SLTRGHSAVRLVYVLEALTNFLNHRITPIVPRGSIASGDLSPSTYAGATGHPDVYKH 240

Db 175 SLTRGHSAVRLVYVLEALTNFLNHRITPIVPRGSIASGDLSPSTYAGATGHPDVYKH 234

Qy 241 VHEGTEKIMPARAISTFGLAAYVLGPKKGLGVNGTAVSASVATLSLDSHMLSLSQ 300

Db 235 VHEGKEKILYAREBAMLFLEPVVLGPKKGLGVNGTAVSASVATLSLDSHMLSLSQ 294

Qy 301 ALTALTEAMVGOOGSPAFPIHDVCRPHPGQVEVARINIRTLSSGSPAVEHEEVKXKD 360

Db 295 SLTAMTEAMVGHAGSFHPIHDVTRPHPTQIEVAGNIRKLEGSRFVAHHEEVKXKD 354

Qy 361 EGIIRQDRYPLRTSPQIGPLVEDMMAVSTLSLE-NNTTNDNPLLDVENKOTAHGNGFQ 419

Db 355 EGIIRQDRYPLRTSPQIGPLVEDMMAVSTLSLE-NNTTNDNPLLDVENKOTAHGNGFQ 414

Qy 420 ASAVSISMEKTRIALALIGKLNFTQCTELNMAANRGLPSCLAEDPSLNYHGKGLDIHI 479

Db 415 AAAYANTMEKTRIGLAQIGKLNFTQCTELNMAANRGLPSCLAEDPSLNYHGKGLDIHI 474

Qy 480 AAYSELGHLANPVTTTPVQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539

Db 475 AAYSELGHLANPVTTTPVQPAEMNGQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 534

Qy 540 VDLRAEMLDFKQDPFLPTLLQCHLG---TGLDV-NALALEVKAANKRLBQTTDYLE 595

Db 535 IDLRAXEPEFKQGFPAIVSLIDQFGSAMTGSNLDELVEKVKTKLAKRLBQTNISDYLV 594

Qy 596 PRMHDAFSYATGTVELLSSSPSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSSQ 655

Db 595 PRMHDAFSYATGTVELLSSSPSANVTTLTAVNAKVASAEKALSLTREVRNRFWQTPSSQ 651

Qy 656 APAAHYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKDGRIHVL 715

Db 652 SPALSYLSPTRTVLVSFVREELGVQARRGDVFGVQOETIGSNVRIYEAIKDGRIHVL 711

Qy 716 VKMLA 720

Db 712 LKMLA 716

RESULT 10

US-10-476-193-21

Sequence 21, Application US/10476193

GENERAL INFORMATION:

APPLICANT: Saitasian, Sima

APPLICANT: Tang, Xiao-Song

APPLICANT: Qi, Wei Wei

TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

FILE REFERENCE: C11777

CURRENT APPLICATION NUMBER: US/10/476,193

CURRENT FILING DATE: 2003-10-28

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

SEQ ID NO 21

LENGTH: 716

TYPE: PRT

ORGANISM: Rhodotorula glutinis

FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (149)..(149)
OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (202)..(202)
OTHER INFORMATION: X=Val, Met, Leu, Cys
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540)..(540)
OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540)..(540)
OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-193-21

Query Match 72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2.3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

1 MAPSLDLSLTTLANGTNGSHAPTKSAAGPTSAARTRPGDGHAAHQSGLIYQELISD 60
1 MAPSLDISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTDIVERKMLAA 54
61 PTDVVELSGYSLTVRDVVGAAKRGRRVONDDIRARVDKSDVDFLKAQONSYYGVT 120
55 PDDSTLELDGSLNLDGVSAARKGRPRVVKSDIRSKIDKSVFRLSGLSMGYGVT 114
121 GFGGSAADTTEDAVSIQKALIHQICGVTPTXSFSVGRGIENTLPLEVVGAMVIRVN 180
115 GFGGSAADTTEDAVSIQKALIEHQICGVLPSFDFRLGRLGLENLPLEVVRGAMTIRVN 174
181 SLTRGSAARLVVLEALTNFLNHRITPIVPLRGSSASGDSLPSYIAGATTGHPDVVH 240
175 SLTRGSAARLVVLEALTNFLNHRITPIVPLRGSSASGDSLPSYIAGATTGHPDVVH 234
241 VLHEGTEKIMFAREALSLFGLBAVVLGPKGGLGVNGTAVASNAATSLDHSMLSLISQ 300
235 VVHEGKEKILVAREAMALFNLEPVVLGPKGGLGVNGTAVASNAATSLDHSMLSLISQ 294
301 ALTALVVEAMVQOQSFAPFIHDVCRPHGQVEVARNRTLLSSGSPFVHEEVEKYVD 360
295 SLTAMTVEAMVQAGSFHFLHDVTRPHPTQTEVAGNIRKLEGGSPFVHHEEVEKYVD 354
361 EGI LRDRYPLRTSPQGLPVEDMHAYSTLSLE-NTTTDNPLLDVENKQTAAHGNFQ 419
355 EGI LRDRYPLRTSPQGLPVEDMHAYSTLSLE-NTTTDNPLLDVENKQTAAHGNFQ 414
420 ASAVSISMEXTRLALALIGKLNFTQCTELLNPAANRGCLPSCLAADPBLNTHGKGLDHI 479
415 AAAVANTMEKTRLGIAQIGKLNFTQCTELLNPAANRGCLPSCLAADPBLNTHGKGLDIAA 474
480 AAASVSELGHANVTTFVCPAENQAVNSLALISARPTAEANDVLSLLASHLYCTIOA 539
475 AATSEIGHANVTTHVQPAEMANOAVNSLALISARPTAEANDVLSLLASHLYCTIOA 534
540 VDLRAVELDKQFDLPTLLQOHLG--TGALDY-NALALEVKKALNKRLEQTTVYLE 595
535 IDLRAXEFBKQFGPAIVSLIDQHFSGAMTGSNLRDELVEKVNITLAKRLEQTNVYLV 594
596 PRMHDAFSYATGVVVLSSSPAVNTLAVNAAKVAESAIAISTREVRNRFQOTPSQ 655
595 PRMHDAFSYATGVVVLSSSPAVNTLAVNAAKVAESAIAISTREVRNRFQOTPSQ 651
656 APAAAYSPSTRVLYFVFEELGVCARRGDVVGVOQETIGSNVRIEATIKGGINVL 715
652 SPLASTISPTQTLVAFVEELGVCARRGDVVGVOQETIGSNVRIEATIKGGINVL 711
716 VKMLA 720
712 LKMLA 716

RESULT 11
US-10-476-198-20
Sequence 20, Application US/10476198
GENERAL INFORMATION:
APPLICANT: Sarsiani, Sina
APPLICANT: Tang, Xiao-Song
APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
FILE REFERENCE: CL177
CURRENT APPLICATION NUMBER: US/10/476.198
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 20
LENGTH: 716
TYPE: PRM
ORGANISM: Rhodotorula glutinis
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (181)..(181)
OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (235)..(235)
OTHER INFORMATION: X=Ala, Gly, Ser, Thr, Pro
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540)..(540)
OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540)..(540)
OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-20

Query Match 72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2.3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

1 MAPSLDLSLTTLANGTNGSHAPTKSAAGPTSAARTRPGDGHAAHQSGLIYQELISD 60
1 MAPSLDISHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTDIVERKMLAA 54
61 PTDVVELSGYSLTVRDVVGAAKRGRRVONDDIRARVDKSDVDFLKAQONSYYGVT 120
55 PDDSTLELDGSLNLDGVSAARKGRPRVVKSDIRSKIDKSVFRLSGLSMGYGVT 114
121 GFGGSAADTTEDAVSIQKALIHQICGVTPTXSFSVGRGIENTLPLEVVGAMVIRVN 180
115 GFGGSAADTTEDAVSIQKALIEHQICGVLPSFDFRLGRLGLENLPLEVVRGAMTIRVN 174
181 SLTRGSAARLVVLEALTNFLNHRITPIVPLRGSSASGDSLPSYIAGATTGHPDVVH 240
175 SLTRGSAARLVVLEALTNFLNHRITPIVPLRGSSASGDSLPSYIAGATTGHPDVVH 234
241 VLHEGTEKIMFAREALSLFGLBAVVLGPKGGLGVNGTAVASNAATSLDHSMLSLISQ 300
235 VVHEGKEKILVAREAMALFNLEPVVLGPKGGLGVNGTAVASNAATSLDHSMLSLISQ 294
301 ALTALVVEAMVQOQSFAPFIHDVCRPHGQVEVARNRTLLSSGSPFVHEEVEKYVD 360
295 SLTAMTVEAMVQAGSFHFLHDVTRPHPTQTEVAGNIRKLEGGSPFVHHEEVEKYVD 354
361 EGI LRDRYPLRTSPQGLPVEDMHAYSTLSLE-NTTTDNPLLDVENKQTAAHGNFQ 419
355 EGI LRDRYPLRTSPQGLPVEDMHAYSTLSLE-NTTTDNPLLDVENKQTAAHGNFQ 414
420 ASAVSISMEXTRLALALIGKLNFTQCTELLNPAANRGCLPSCLAADPBLNTHGKGLDHI 479
415 AAAVANTMEKTRLGIAQIGKLNFTQCTELLNPAANRGCLPSCLAADPBLNTHGKGLDIAA 474

```

QY 480 AAYASELGHANPVTTFVQPAENQAVNSLALISARPTAEANDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHANPVTTHVQPAEMANOAVNSLALISARPTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAEMLDPKKQDPPLPTLLQOHLG---TGLDV-NALALEVKKALNKRLQCTTYLE 595
DB 535 IDLRAXEPEFKKQFPALVSLIDQHFSGAMTGSNLDLVEKVNKTLLAKLEQTNLYDLV 594
QY 596 PRMHDAFSYATGIVVELLSSSPSANVTLLAVANAKVASAEKALISLREVRNRFMQTPSSQ 655
DB 595 PRMHDAFSYATGIVVELLSSSPSANVTLLAVANAKVASAEKALISLREVRNRFMQTPSSQ 651
QY 656 APAAVLSPTRTVLYSFRRELGVARRGDVPFVQOQETIGSNVRIYEAIXDGRINVL 715
DB 652 SPALSYLSPRTQILYAFREELGVAKRGDVFILGQEVTTIGSNVSKTYEAIKSGRINVL 711
QY 716 VKMLA 720
DB 712 VKMLA 716

RESULT 12
US-10-476-198-21
; Sequence 21, Application US/10476198
; GENERAL INFORMATION:
; APPLICANT: Sariafian, Sima
; APPLICANT: Tang, Xiao-song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,198
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X=Pro, Ala, Ser, Thr, Gly
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X=Val, Met, Leu, Cys
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly
US-10-476-198-21

Query Match 72.1%; Score 2603; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 2,3e-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

QY 1 MASELSLATTLANGSTNGSHAPPTSAAGPTSAIRPTPGDLGHAHQSLEIYOELLSD 60
DB 1 MASELSLSHSFANG-----VASAKQAVNGASTNLAVAGSHLPTTQVTVDIYVKMLAA 54
QY 61 PTDDVVELSGYSTLVEDVGAARKGRVAVQNDDEIRAVDSQVDFLKAQOLNSVYGVTT 120
DB 55 PTGSTLELDGYSLNLDQVWSAARKGRPVAVKQSDERIKIDSVFELISQLSMEVYGVTT 114
QY 121 GFGGSAADRTEDAVSLQAKLIEHQLCGVTPTSXSFVSGRGLENTLPLEVVRGAMVIRVN 180
DB 115 GFGGSAADRTEDAVSLQAKLIEHQLCGVLPSSFXDFRLGRGLENSLPLEVVRGAMVIRVN 174

```

```

QY 181 SLTRGSAVRLVLEALNFINLNRITPIVPIRGSSISASGDISPLSYAGATGHPDVKVH 240
DB 175 SLTRGSAVRLVLEALNFINLNRITPIVPIRGSSISASGDISPLSYIAAAISGHPDBKVH 234
QY 241 VHEGTEKIMEAREALISFGLEAVYLGRKEGIGVNGVAVASNAWATSLDHSHTLSLSQ 300
DB 235 VHEGTEKILYAREALNLEPVVLGPKBGLGVNGTAVASNAWATSLDHSHTLSLSQ 294
QY 301 ALTALTYEAMVQOQSFAPFTHDYCRPHQGVVEYARNIRTLSSGSEFAVEHEEVKQKD 360
DB 295 SLTAMTYEAMVQOQSFAPFTHDYCRPHQGVVEYARNIRTLSSGSEFAVEHEEVKQKD 354
QY 361 EGLHQDRYPLFTSQPGLVEDMMAHXYLSLE-NTTTPDNLLDVENKQTAHGNFQ 419
DB 355 EGLHQDRYPLFTSQPGLVEDMMAHXYLSLE-NTTTPDNLLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKRLALALIGKINFTQCTELNNAANRGIPSCIAADEPSLNYHGKGLDIH 479
DB 415 AAVANTMEKRLGALQIGKINFTQCTELNNAANRGIPSCIAADEPSLNYHGKGLDIH 474
QY 480 AAYASELGHANPVTTFVQPAENQAVNSLALISARPTAEANDVLSLLASHLYCTLOA 539
DB 475 AAYTSELGHANPVTTHVQPAEMANOAVNSLALISARPTESNDVLSLLATHLYCVLOA 534
QY 540 VDLRAEMLDPKKQDPPLPTLLQOHLG---TGLDV-NALALEVKKALNKRLQCTTYLE 595
DB 535 IDLRAXEPEFKKQFPALVSLIDQHFSGAMTGSNLDLVEKVNKTLLAKLEQTNLYDLV 594
QY 596 PRMHDAFSYATGIVVELLSSSPSANVTLLAVANAKVASAEKALISLREVRNRFMQTPSSQ 655
DB 595 PRMHDAFSYATGIVVELLSSSPSANVTLLAVANAKVASAEKALISLREVRNRFMQTPSSQ 651
QY 656 APAAVLSPTRTVLYSFRRELGVARRGDVPFVQOQETIGSNVRIYEAIXDGRINVL 715
DB 652 SPALSYLSPRTQILYAFREELGVAKRGDVFILGQEVTTIGSNVSKTYEAIKSGRINVL 711
QY 716 VKMLA 720
DB 712 VKMLA 716

RESULT 13
US-10-476-193-18
; Sequence 18, Application US/10476193
; GENERAL INFORMATION:
; APPLICANT: Sariafian, Sima
; APPLICANT: Tang, Xiao-song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/476,193
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X=Gly, Ala, Ser, Thr
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X=Leu, Met, Ile, Val, Cys
; NAME/KEY: MISC FEATURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-193-18

```


Query Match 72.0%; Score 2599; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 4,76-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

QY 1 MAPSLDLSATTLANGFTNGSHAAPKSAAGPTSA LRRTPGLDGHAHQSOLEIYQELISD 60
Db 1 MAPSLDLSISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTQTQYVDIYEKLLAA 54
QY 61 PDDVVELSGYSLIVRDYVGARKGRVRVQNDDEIRAVDKSVDFLKAQIQNSVGYTT 120
Db 55 PDDSTLELDGYSILNGDVSAARKGRPVKVDSDERKIDKSVDFRSQLSMSVGYTT 114
QY 121 GEGSADRTTEAVSLOKALIEHQICGVPTXSFSFGRENTLPLEVVGAMVIRVN 180
Db 115 GEGSADRTTEKALISQKALLEHLCVLPSSFDSPFRGRLENSLPLEVVGAMVIRVN 174
QY 181 SLTRGSAVRVLVLEALNFINHRITPIVPLRGSIASAGDLSPLSYIAGAITGHPDKVH 240
Db 175 SLTRGSAVRVLVLEALNFINHGITPIVPLRGITISAGDLSPLSYIAAISGHPDKVH 234
QY 241 VHEGHEKIMPARBAISLFGLEAVVLGPKKGLGVNGTAVASAMATLISHSHLISLSQ 300
Db 235 VHEGHEKILYAREMAALFNLEPVVLGPKKGLGVNGTAVASAMATLISHSHLISLSQ 294
QY 301 ALTALTVEMVQOQGSFAPFIHDVCRPHPGQVEVARNIRTLISGSSFAVEHEEVRKYKD 360
Db 295 SLTAMTVEMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKLLGSRFAVHHEEVRKYKD 354
QY 361 EGI LRDRPLRTSPQFLGPIVEDMHAYSTLSLE-NTTTDNPFLDVENKQTAHGNFQ 419
Db 355 EGI LRDRPLRTSPQFLGPIVEDMHAYSTLSLE-NTTTDNPFLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNANMRGLPSCIAEDPSLNYHKGDLIHI 479
Db 415 AAAVANTMEKTRIGLAQIGKLNFTQCTELNANMRGLPSCIAEDPSLNYHKGDLIHI 474
QY 480 AAVASLGHIANPVTTFPQAPAMQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAVTSELGHIANPVTTFPQAPAMQAVNSLALISARRTESNDVLSLLATHLYCYLOA 534
QY 540 VDLRAMELDFKKQFDPPLPTLIQOHLG---TGLDV-NALALEVKKALKLEQTTTYDLE 595
Db 535 IDLRAXEFPFKQFQPAIVSLIDQHFSGAMTGSNLRDELVEKVKTKLAKLEQNSYDLV 594
QY 596 PRMHDAFSYATGVVELSSSPSANVTLTAVANMKVSAEKAISLTREVRNRFMQTSQ 655
Db 595 PRMHDAFSYATGVVELSSSPSANVTLTAVANMKVSAEKAISLTREVRNRFMQTSQ 651
QY 656 APHAHLSRTVLYSVFVEELGVQARRGDVFGVQOCTIGSNVRIYEAIKQRIHNVL 715
Db 652 SPALSTLSPTQILYAFVFEELGVQARRGDVFGVQOCTIGSNVRIYEAIKSGRINNVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 14
US-10-476-198-18
Sequence 18, Application US/10476198
GENERAL INFORMATION:
APPLICANT: Sariaeiani, Sima
APPLICANT: Tang, Xiao-Song
APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
FILE REFERENCE: CU1777
CURRENT APPLICATION NUMBER: US/10/476,198
CURRENT FILING DATE: 2003-10-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 18

LENGTH: 716
TYPE: PRT
ORGANISM: Rhodocorula glutinis
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (126) ..(126)
OTHER INFORMATION: X=Gly, Ala, Ser, Thr
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (138) ..(138)
OTHER INFORMATION: X=Leu, Met, Ile, Val, Cys
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (540) ..(540)
OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly
US-10-476-198-18

Query Match 72.0%; Score 2599; DB 6; Length 716;
Best Local Similarity 71.7%; Pred. No. 4,76-193;
Matches 520; Conservative 85; Mismatches 106; Indels 14; Gaps 5;

QY 1 MAPSLDLSATTLANGFTNGSHAAPKSAAGPTSA LRRTPGLDGHAHQSOLEIYQELISD 60
Db 1 MAPSLDLSISHSPANG-----VASAKQAVNGASTNLAVAGSHLPTQTQYVDIYEKLLAA 54
QY 61 PDDVVELSGYSLIVRDYVGARKGRVRVQNDDEIRAVDKSVDFLKAQIQNSVGYTT 120
Db 55 PDDSTLELDGYSILNGDVSAARKGRPVKVDSDERKIDKSVDFRSQLSMSVGYTT 114
QY 121 GEGSADRTTEAVSLOKALIEHQICGVPTXSFSFGRENTLPLEVVGAMVIRVN 180
Db 115 GEGSADRTTEKALISQKALLEHLCVLPSSFDSPFRGRLENSLPLEVVGAMVIRVN 174
QY 181 SLTRGSAVRVLVLEALNFINHRITPIVPLRGSIASAGDLSPLSYIAGAITGHPDKVH 240
Db 175 SLTRGSAVRVLVLEALNFINHGITPIVPLRGITISAGDLSPLSYIAAISGHPDKVH 234
QY 241 VHEGHEKIMPARBAISLFGLEAVVLGPKKGLGVNGTAVASAMATLISHSHLISLSQ 300
Db 235 VHEGHEKILYAREMAALFNLEPVVLGPKKGLGVNGTAVASAMATLISHSHLISLSQ 294
QY 301 ALTALTVEMVQOQGSFAPFIHDVCRPHPGQVEVARNIRTLISGSSFAVEHEEVRKYKD 360
Db 295 SLTAMTVEMVGHAGSFHFLHDVTRPHPTQIEVAGNIRKLLGSRFAVHHEEVRKYKD 354
QY 361 EGI LRDRPLRTSPQFLGPIVEDMHAYSTLSLE-NTTTDNPFLDVENKQTAHGNFQ 419
Db 355 EGI LRDRPLRTSPQFLGPIVEDMHAYSTLSLE-NTTTDNPFLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELNANMRGLPSCIAEDPSLNYHKGDLIHI 479
Db 415 AAAVANTMEKTRIGLAQIGKLNFTQCTELNANMRGLPSCIAEDPSLNYHKGDLIHI 474
QY 480 AAVASLGHIANPVTTFPQAPAMQAVNSLALISARRTAEANDVLSLLASHLYCTLOA 539
Db 475 AAVTSELGHIANPVTTFPQAPAMQAVNSLALISARRTESNDVLSLLATHLYCYLOA 534
QY 540 VDLRAMELDFKKQFDPPLPTLIQOHLG---TGLDV-NALALEVKKALKLEQTTTYDLE 595
Db 535 IDLRAXEFPFKQFQPAIVSLIDQHFSGAMTGSNLRDELVEKVKTKLAKLEQNSYDLV 594
QY 596 PRMHDAFSYATGVVELSSSPSANVTLTAVANMKVSAEKAISLTREVRNRFMQTSQ 655
Db 595 PRMHDAFSYATGVVELSSSPSANVTLTAVANMKVSAEKAISLTREVRNRFMQTSQ 651
QY 656 APHAHLSRTVLYSVFVEELGVQARRGDVFGVQOCTIGSNVRIYEAIKQRIHNVL 715
Db 652 SPALSTLSPTQILYAFVFEELGVQARRGDVFGVQOCTIGSNVRIYEAIKSGRINNVL 711
QY 716 VKMLA 720
Db 712 LKMLA 716

RESULT 15
US-10-476-193-17
Sequence 17, Application US/10476193
GENERAL INFORMATION:
APPLICANT: Sartiasani, Sima
APPLICANT: Tang, Xiao-song
APPLICANT: Qi, Wei wei
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
FILE REFERENCE: C11777
CURRENT APPLICATION NUMBER: US/10/476,193
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 17
LENGTH: 716
TYPE: PRT
ORGANISM: Rhodocorula glutinis
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (126)..(126)
OTHER INFORMATION: X= Gly, Ser, Ala, Thr
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (138)..(138)
OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (149)..(149)
OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (181)..(181)
OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (198)..(198)
OTHER INFORMATION: X= Asp, Asn, Glu, Gln
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (202)..(202)
OTHER INFORMATION: X= Val, Met, Leu, Cys
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (235)..(235)
OTHER INFORMATION: X= Ala, Gly, Ser, Pro
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (502)..(502)
OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (540)..(540)
OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-10-476-193-17

Query Match 71.4%; Score 2579; DB 6; Length 716;
Best Local Similarity 71.0%; Pred. No. 1,7e-191;
Matches 515; Conservative 86; Mismatches 110; Indels 14; Gaps 5;

QY 1 MASELSLATTLLANGFTNSHAAPTSAAGPTSAARTRTPGDLGHAHSGLEIVQELISD 60
DB 1 MASELSLISHSIFANG-----VASAKQAVNGASTNLAVAGSHLPTTYQTVQVIVKMLAA 54
QY 61 PTDDVVELSGSYLTVDVVGAAKGRVAVQNDDEIRARVDKSVDFLKAQOLONSIVGYTT 120
DB 55 PTOSTLELDGYSLNLDGVVSAARKGRPVAVKDSDEIRSKIDKSVDFLRSQLSMSVGYTT 114
QY 121 GFGGSADTRDEDVSLQKALIEHQLCGVTPPTXSXSFVGRGIENTLPLEVTRGAMVTRVN 180
DB 115 GFGGSADTRTEXAISIQKALIEHQLCGVLPSSFDXFLGRGLENSLPLEVTRGAMTIRVN 174

QY 181 SLTRGSHAVRLVLEALTNFNLNHRITPIVPLRGSISASGDISPLSYAGAITGHPDVKH 240
DB 175 SLTRGSHAVRLVLEALTNFNLNHRITPIVPLRGTISASGDISPLSYAAISGHPDSKVH 234
QY 241 VHEGTEKIMFARBAISLFGLEAVVLGPKEGELGVNGTAVASAATSLSDSHLSLSLQ 300
DB 235 XHBEKEKILYREMAALPNEPVLGPKEGELGVNGTAVASAATSLDHDAMLSLSLQ 294
QY 301 ALTALTEAMVQGGSPAFPHDVCRRPHGQVEVARNIRTLSSGSPAVHEEVEVKYKD 360
DB 295 SLTAMTEAMVGHAGSPFPHLDVTRPAPTQIEVAGNIRKLSSGSPAVHHEEVEVKYKD 354
QY 361 EGLRQDRYPLRTPSPQIFGLVEDMHAYSLSLE-NNTTDNPLDVENKQTAHGNFQ 419
DB 355 EGLRQDRYPLRTPSPQIFGLVEDMHAYSLSLE-NNTTDNPLDVENKQTAHGNFQ 414
QY 420 ASAVSISMEKTRIALALIGKLNFTQCTELLNAAMNRGLPSCLAEDPSLNYHGKGLDHI 479
DB 415 AAAVANTWEKTRIGLAQIGKLNFTQCTELMANAGNRGLPSCLAEDPSLSYHGKGLDIAA 474
QY 480 AAYASELGHANPVTTPVQPAEMNQAVNSIALISARTAEANDVLSLLASHYCTLOA 539
DB 475 AAYTSELGHANPVTTHVQPAEMNQAVNSIALISARTTESNDVLSLLATHYCYLOA 534
QY 540 VDLRAVELDPKKQFDPPLPTLLOQHTG--TGLDV-NALALEYKALNKLEQTTVDLE 595
DB 535 IDLRALIEPFFKQFGPAPALVSLDQHFSGAMTGSULREIYENKNTLAKLEQTNISYDV 594
QY 596 PRMHDAFSYATGVVELLSSPSSANVTITAVANAKVASAEKALSTREVENRFTWQTPSSQ 655
DB 595 PRMHDAFSYATGVVELLSSPSSANVTITAVANAKVASAEKALSTREVENRFTWQTPSSQ 651
QY 656 APAHAAYLSPTRVLYSFPREELGYQARRGDVFGVQOETIGSNVSRITYEAIKGRINHYL 715
DB 652 SPALSYLSERTOILYAFVREELGYQARRGDVFLGKQEVITIGSNVSKITYEAIKGRINNYL 711
QY 716 VKXLA 720
DB 712 LKXLA 716

Search completed: September 9, 2004, 10:21:12
Job time : 46.8174 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:58:38 ; Search time 28.8797 Seconds
(without alignment)
2398.154 Million cell updates/sec

Title: US-09-939-408a-13
Perfect score: 3610
Sequence: 1 MAPSLDLSLATTIANGFTNGS.....RIYKAIKGRINHVLVKMLA 720

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 78:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2614	72.4	716	2	A56628 phenylalanine ammo
2	2573	71.3	693	2	A29607 phenylalanine ammo
3	2499.5	69.2	713	2	S01899 phenylalanine ammo
4	1041.5	28.9	725	2	G84787 phenylalanine ammo
5	1035.5	28.7	725	2	S52990 phenylalanine ammo
6	1017.5	28.2	718	2	S48726 phenylalanine ammo
7	1016.5	28.2	715	2	S66343 phenylalanine ammo
8	1016	28.1	715	2	J02265 phenylalanine ammo
9	1015.5	28.1	717	2	T46172 phenylalanine ammo
10	1011	28.0	713	2	S22891 phenylalanine ammo
11	1009	28.0	725	2	S17444 phenylalanine ammo
12	1006	27.9	716	2	S04463 phenylalanine ammo
13	1005.5	27.9	723	2	S25303 phenylalanine ammo
14	1005	27.8	712	2	T03663 phenylalanine ammo
15	1005	27.8	716	2	S48725 phenylalanine ammo
16	1004	27.8	701	2	S06475 phenylalanine ammo
17	1003.5	27.8	712	2	S04127 phenylalanine ammo
18	998	27.6	705	2	J05873 phenylalanine ammo
19	997	27.6	720	2	S21174 phenylalanine ammo
20	995	27.6	710	2	J05872 phenylalanine ammo
21	994.5	27.5	721	2	A44133 phenylalanine ammo
22	993.5	27.5	708	2	T10909 phenylalanine ammo
23	993	27.5	712	2	T01858 phenylalanine ammo
24	987.5	27.4	667	2	T12749 phenylalanine ammo
25	983	27.2	710	2	T14295 phenylalanine ammo
26	982.5	27.2	708	2	S66313 phenylalanine ammo
27	973.5	27.0	700	2	T06545 phenylalanine ammo
28	973.5	27.0	754	2	T09777 phenylalanine ammo
29	966	26.8	550	2	S70916 phenylalanine ammo

30	964.5	26.7	695	2	S52992 phenylalanine ammo
31	955	26.5	549	2	T05970 phenylalanine ammo
32	954.5	26.4	707	2	S29029 phenylalanine ammo
33	953.5	26.4	710	2	S60042 phenylalanine ammo
34	946.5	26.2	710	2	S04128 phenylalanine ammo
35	946	26.2	619	2	S52632 phenylalanine ammo
36	934	25.9	716	2	S28185 phenylalanine ammo
37	875.5	24.3	582	2	S60043 phenylalanine ammo
38	703.5	19.5	506	2	S04129 phenylalanine ammo
39	697.5	19.3	505	2	A24727 phenylalanine ammo
40	659.5	18.3	497	2	T05968 probable phenylala
41	655.5	18.2	475	2	T05966 phenylalanine ammo
42	547.5	15.2	524	2	T75610 histidine ammonia-
43	543.5	15.1	416	2	T01070 phenylalanine ammo
44	543	15.0	514	2	A63040 histidine ammonia-
45	543	15.0	514	2	E98245 huth gene homolog

ALIGNMENTS

RESULT 1

A56628 phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodospiridium toruloides) (strain CB C:Species: Rhodospiridium toruloides
C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C:Accession: A56628
R:Kasumussen, O.F.; Oerum, H.
DNA Seq. 1, 207-211, 1991
A:Title: Analysis of the gene for phenylalanine ammonia-lyase from Rhodospiridium torulo
A:Reference number: A56628; MUID:92127061; PMID:1773059
A:Accession: A56628
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-716 <RAS>
A:Cross-references: GB:X51513; NID:g3293; PIDN:CA3586.1; PID:g3294
A:Note: sequence extracted from NCBI Backbone (NCBI:78381)
C:Genetics:
A:Gene: pal
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:211-213/Cross-link: 5-Imlidazolinone (Ala-Gly) #status predicted
F:212/Modified site: dehydroalanine (Ser) #status predicted

Query Match	72.4%;	Score 2614;	DB 2;	Length 716;
Best Local Similarity	72.0%;	Pred. No. 1.3e+165;		
Matches 522;	Conservative 86;	Mismatches 103;	Indels 14;	Gaps 5;
QY	1	MAPSLDLSLATTIANGFTNGSHPAFTKSAAGPTGALRTTGLDGHAGSQLEIYQELISD 60		
DB	1	MAPSLDLSLHSHFANG-----VASAKQAVNGASTMLAVAGSHLPTTQVTVQDIYKMLA 54		
QY	61	PPDDVVEISGYSLTTRDVVGARKGRRYVONDDRTBARVDKSVDFLKAQVQNSYGVTT 120		
DB	55	PTDSTLELDGYSLNTGDDVVSARKGRPVKXDSDEIRSKIDKVSFFLSQLSMAYGVTT 114		
QY	121	GFGSADTETEDAVSLQKALIEHQCGVTPTSXSSFSVGRGLENTPLEVVYRGAMVIRVN 180		
DB	115	GFGSADTETEDALISQKALIEHQCGVLPSSFDFFRGHLENPLPVEVYRGAMTIRVN 174		
QY	181	SLTRGSAVRVYVLEALTNFLNRITPIVPLRGSSISAGDLSPISTIGAGTATGHPDVYH 240		
DB	175	SLTRGSAVRVYVLEALTNFLNRITPIVPLRGTSISAGDLSPISTIYAAASGHPDSKVH 234		
QY	241	VHEGTEKIMFARERASLFGLEBAVLGPEKGLGVNGAVASAMATLSLHSHMLSLLSQ 300		
DB	235	VHEGTEKILYAREMALFNLEPVLGPEKGLGVNGAVASAMATLALHNAHMLSLLSQ 294		
QY	301	ALTALTEAMVQCGQSPAFPIHDVCRPHPGQVEVARNIRTLISGSSPFAVHEBEYKVD 360		
DB	295	SLTAMTEAMVQAGSFHPIHLDVTRPHTQIEVAGNIRKLLGSRFAVHHEBEYKVD 354		
QY	361	EGILRQDRPLRTSPQFLGPIVEDMMHAYSTLSLE-NNTTTNDPLDVENVQTAHGQNFQ 419		

```

Db      355  EGILRQDRYPTRTSHQMLGPIVSDLIHAAHVAITLHAGSTIDNPLIDENKTTSHHGQNFQ  414
Qy      420  ASAASISNEKTRLALALIGKLNFTQCTELNAAANNRGLPSCIAADPESLNYHGKGLDIHI  479
Db      415  AAAVANMTEKTRLGIAQIGKLNFTQITLMLNAAANNRGLPSCIAADPESLNYHGKGLDIAA  474
Qy      480  AAAYSEIGHLNPWTTFVQPAEMGNOANSIALTSARETAANVLTSLLSASHYCTLOA  539
Db      475  AAAYSEIGHLANPVTHVQPAEMANOAVNSIALTSARETTSDNLSLLTLHIVCVLOA  534
Qy      540  VDLRAMELDFKKQEDPLPLTLLOOHTG--TGLDV-NALALEYVKALKRLEQTTTYDLE  595
Db      535  IDLRRIEPEFKQGFQCPAIVSLIDQHPGSAMTGSNLRDELVEKVKVTKLAKRIEQTNSYDV  594
Qy      596  PRMHDAPEYAGTVEVELSSPSPANVTLTANNAKMYASAEKALISTREYVRIRKFPQTPESQ  655
Db      595  PRMHDAPEFAAGTVEVELST--SLSLAANAMVMAAEASALSTRQVRETFWSAASTS  651
Qy      656  AAPAAAYLSPTRLVYSFVREELGYQARRGDFVFGVOQETISNSYRIEATIKDIRINHVLT  715
Db      652  SPALSYLSPRQIIYAFVREELGYKARRGDFVFGKQEVTLISNSYKITEALIKSIRINNVLT  711
Qy      716  VKMLA 720
Db      712  LKMLA 716

```

```

RESULT 2
A29607
phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodosporidium toruloides) (strain IM-1)
C:Species: Rhodosporidium toruloides
C:Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 16-Jul-1999
C:Accession: A29607
R:Anson, J.G.; Gilbert, H.J.; Oram, J.D.; Manton, N.P.
Gene 58, 189-199, 1987
A:Title: Complete nucleotide sequence of the Rhodosporidium toruloides gene coding for P
A:Reference number: A29607; MUID:88112870; PMID:2828184
A:Accession: A29607
A:Molecule type: DNA
A:Residues: 1-693 <ANS>
A:Cross-references: GB:M8261; NID:g169745; PID:AAA33863.1; PID:g169746
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:188-190/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:189/Modified site: dehydroalanine (Ser) #status predicted

Query Match: 71.3%; Score 2573; DB 2; Length 693;
Best Local Similarity 74.9%; Pred. No. 6.5e-163;
Matches 507; Conservative 83; Mismatches 79; Indels 8; Gaps 4;

```

[illegible]

```

Db      320 VHHBEVKVYDDEGCGILRQRYRPLRTSPQVLGSVSDILHAAVLTTEAGSTTNNPLIDV 375
QY      408 ENKQTAGNGFQASAVSISMEXTRLALIGKLTFTCTELLNAAWNGLPSCUAADPS 467
Db      380 ENKTSHHGNGFQAAAVANTMEKTRBLGAQIGKLTFTQTEMLNGMNRGLPSCUAADPS 439
QY      468 LNHGKGLDHIHAAYSEIGHLANPPTTTFVQAEENGQANVNSIALISKRPAEANDVLSL 527
Db      440 LSHCKGLDIAAAAYNSELGHLANPPTTHQVQAEANQANVNSIALISARPTESNDVLSL 499
QY      528 LLSHXYCTIQAVDILAMELDFKQEDPLPLFTLLQOHLG--TGSLD-NALATEVKKALN 583
Db      500 LLAHLHYCVQALDLALAIFFEFKKQFGPAIVSLIDHFGSAMTGSNLDRLVEYKNTTLA 559
QY      584 KRLQQTITVDLEBRWDAFSYATGTVELLSSPSANVTLLPAAWAKVYASAEKAISLTRE 643
Db      560 KRLQGNNSYDVLVRWMDAFSPAAGTIVEVLST--SLSLAAVNAWKKVAAESAISLTRO 616
QY      644 VNRKFMQTPSSQAPAAAYLSPRTRVLISFVREELQVQRKDDVFGVQOQETISNVSRIT 703
Db      617 VRRTFSSAATSSPAASTYLSPTQILYAFVREELQVQKRDVDLGGQEVITIGSVSKIT 676
QY      704 EAIKDGRIHVLYWKLMA 720
Db      677 EAIKSGRIINVLIKMLA 693

```

RESULT 3
 S01999
 phenylalanine ammonia-lyase (EC 4.3.1.5) - fungus (Rhodocortula rubra)
 C.Species: Rhodocortula rubra
 C.Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
 A.Accession: S01999
 R.Filpula, D.; Vassler, C.A.; Levy, A.; Sykes, A.; Strausberg, R.L.
 Nucleic Acids Res. 16, 11381, 1988
 A.Title: Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodocortula
 A.Reference number: S01999; PMID:89083582; PMID:3205749
 A.Accession: S01999
 A.Molecule type: DNA
 A.Residues: 1-713 <FIL>
 A.Cross-references: EMBL:X13094; NID:G3284; PID:CAA31486.1; PID:G295942
 R.Vassler, C.A.; Strausberg, R.L.; Sykes, A.; Levy, A.; Filpula, D.
 Nucleic Acids Res. 16, 11382, 1988
 A.Title: cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal gen
 A.Reference number: S01990; PMID:89083583; PMID:3205750
 A.Accession: S01990
 A.Status: translation not shown
 A.Molecule type: mRNA
 A.Residues: 1-674 <VAS>
 A.Cross-references: EMBL:X13095
 C.Genetics:
 A.Introns: 110/2; 139/2; 372/3; 428/2; 482/3
 C.Superfamily: histidine ammonia-lyase
 C.Keywords: ammonia-lyase; carbon-nitrogen lyase
 E:217-219/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 I:218/Modified site: dehydroalanine (Ser) #status predicted

```

Query Match:          69.2%; Score 2499.5; DB 2; Length 713;
Best Local Similarity 69.6%; Pred.No.5.2e-158;
Matches 511; Conservative 83; Mismatches 105; Indels 35; Gaps 9.

QY      1 MAPSLDLSATLTANGFTNGSHAAPT-----XSAGPSTALRRTPGLDHAAHOSOLEI 53
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       1 MAPSVDSITSVANSLNYGLHPAAAANGGVNKKTKMAGSLPTT-----ETTOLDI 52

QY      54 VCGELSDP-TDDVAVLSSGS.LTVRDVVGAARKGRRYVRONDEIRARVDVDFPKAOL 112
        ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       53 VERILLADGACATQITLDGYTTLLGDVGGAARGRSKVADSPIREKTIASVEFLRTOLD 112

QY      113 NSVYGVTTFGGSSADTRTEDAVSLQKALLIHQLCGVPTPSXSSFVSGRLENTLPLEVR 172
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       113 NSVYGCTTFGGSSADTRTEDAISIQKLLRHQLCGVPTPSMGFALGRLENLPLEYVR 172

```

QY 173 GAMVIRVNSLTGSHAVLVLEALTNFLNRIPIVPLRGSIASGDLPSLAYIAIT 232
DB 173 GATLTRVNSLTGSHAVLVLEALTNFLNRIPIVPLRGSIASGDLPSLAYIAIT 232
QY 233 GHPDVAVHLHEGTEKINFAREALISFLGEAVLVGPKBGLVNGTAVSASMATLSLHDS 292
DB 233 GHPDASKVHV--DG--KIMGAQALAKGLQPVLCGPKBGLVNGTAVSASMATLTALTD 288
QY 293 HMLSLLSQMLTALTYEAMVGOQGSFAPFIHDVCRHPQGVANRIRLLSGSSFAYEHE 352
DB 289 HVLISLAQALLTALTYEAMVGHAGSFHPLHDVCRHPQGVANRIRLLSGSKYAVHHE 348
QY 353 EEVYKQDEGILRODRYPLRTSPQFLGVLVEDMMHAYSTLSLE--NMTTNDPLDVENKQ 411
DB 349 TEKVAVDDGILRQDRYPLRTSPQFLGVLVEDMMHAYSTLSLENGQSTTDPLDLNKK 408
QY 412 THAGNFOASAVISMEKTRALALIGKLNFTQCTELINAMNRLPSCLAADPPSYNH 471
DB 409 THHGAFPMASVSGNTEKTRIALVALMGKVSFPQLTEMNAGNNALPSCLAADPPSYNH 468
QY 472 GKGLDHIAAVSEIHLANPVTTFVQPAEMGNQAVNSLALISARRTEANDVLSILLAS 531
DB 469 CGKLDIAAAVYSEIHLANPVTTHVQPAEMGNQAVNSLALISARRTEANDVLSILLAT 528
QY 532 HLYCTLOAVDLRAMELDFKKQFDPPLPTLLQOHLGTGLDVNALV-----EYKALKNRL 586
DB 529 HLYCTVLOAVDLRAMELDFKKQFDPPLPTLLQOHLGTGLDVNALV-----EYKALKNRL 582
QY 587 ECTTVDLEPRRHDAFSYATGTVVVELSSPSANVTLTANVAMKVASAKAISITREVN 646
DB 583 QQNNASYDLEQRHDPFSYATGTVVVELSSPSANVTLTANVAMKVASAKAISITREVD 639
QY 647 RFMQTPSSQAPAHAYLSPRTVLYSFVEBELGVQARGDVFVGVOQETISNVSRIYEA 706
DB 640 SFPAANSSSPALKLISPRTRLYSFVREVGVKARQDVLGQGEVITGNVSRIYEA 699
QY 707 KQGRINHVLVKML 720
DB 700 KQGCIAPLVVKMA 713

RESULT 4

G84787
phenylalanine ammonia lyase (PAL) [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C/Accession: G84787
R/Lin: X.; Kaul, S.; Rounseley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Unayam, L.; Tallon, L.;
euse, D.; Niemann, M.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: G84787
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-725 <STO>
A/Cross-references: GB:AEO02093; NID:96598547; PIDN:AAD18156.2; GSPDB:GN00139
C/Genetics:
A:Gene: At2g37040
A:Map position: 2
C:Superfamily: histidine ammonia-lyase
Query Match 28.5%; Score 1041.5; DB 2; Length 725;
Best Local Similarity 36.5%; Pred. No. 3,7e-61;
Matches 272; Conservative 138; Mismatches 258; Indels 77; Gaps 21;
QY 18 NGSNAFPTSAAAGPTGAL-----RTTGGDGHAAHQ---SOLEIVQELISD 60
DB 4 NGAH-----KSNGGGVDAMLCGGDIKTAKNVYINAEPLWAGAAAEQMSKSHLDEVKRYAE 59
QY 61 PTDDVVELSGYSLTVSDVGAARAKGRVYQNDDEIRAVDKSVDFLAKQIQNSV--YGV 118
DB 61 PTDDVVELSGYSLTVSDVGAARAKGRVYQNDDEIRAVDKSVDFLAKQIQNSV--YGV 118

RESULT 5

S52990
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1999
C/Accession: S52990; PQ0140
R/Werner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
Plant Mol. Biol. 27, 327-338, 1995
A/Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A/Reference number: S52990; MUID:95195160; PMID:7886622
A/Accession: S52990
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-725 <WAN>
A/Cross-references: EMBL:L33677; NID:9497418; PIDN:AAC18870.1; PID:9497419
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
Plant Cell 2, 837-848, 1990
A/Title: Functional properties of a phenylalanine ammonia-lyase promoter from Arabidopsis
A/Reference number: PQ0140; MUID:93005677; PMID:2152131
A/Accession: PQ0140
A/Molecule type: DNA
A/Residues: 1-240 <OHL>
A/Cross-references: GB:X62747; NID:916408; PIDN:CAA44609.1; PID:916409
A/Experimental source: Ecotype Columbia
C/Comment: phenylalanine ammonia-lyase catalyzes the deamination of L-phenylalanine to t

C:Genetics:
A:Gene: PAL1
A:Introns: 143/2
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:212-213/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F:212/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.7%; Score 1035.5; DB 2; Length 725;
Best Local Similarity 36.4%; Pred. No. 9.2e-61;
Matches 271; Conservative 138; Mismatches 259; Indels 77; Gaps 21;

```
QY 18 NGSHAPTKSAGTSL-----RRTGGLGHAHQ---SOLEYQELSD 60
DB 4 NGAH---KSNGGVDNMLCGGDIKTKNNTVNAEDBLNKGAAAEQKSHLDEVKRYAE 59
QY 61 PTDDVVELSGYSLTVADVGAARKGRVQNDDEIRARVDKSVDFLKAQLQNSV--YGV 118
DB 60 FRKRVNVLGGETLLTGVAALSTIGKSVKVELSETARAGVNASDWMVSMKKTDSYGV 119
QY 119 TTGGGADTRTEDANSLOKALIEHQLCGVTPSKSSFSVCGLENTLPLEVNGAMVIR 178
DB 120 TTGGGATSHRTKNGALOELRLNAGI-----FQSTKETSHTLPHSATRAMMIVR 172
QY 179 VNSLTRGSAVRLVLEALTNFNLNHRITPVLPGSISASGDLSPSYAGATGHPVK 238
DB 173 INTLQFGSGIRFEILEITISFLNNITPSLPRGTTASGDLVLSYIAGILGRPSK 232
QY 239 VHVHBSGE-KIMPAEALSLFGLA--VVLGPKGGLVNGTAVASAMATLSLHDSML 295
DB 233 A---TGPNGEALTLAEAFKLAGISSGFEDLPQKGLALVNGTAVGSGMAGVVFETIVL 288
QY 296 SLTSLATLTALVAMVGGQGSFAPFIDVCRPHGQVEVARNIRTLSSGSAVEHEEV 355
DB 289 SVLAEILSAVFAEWGKRP-EFTDHLTHLKHHPQCEIAAAMEHILDGSSY-MKLAOKL 346
QY 356 KVXDDGILRQDRYPLRTSPQFLGVLVEDMMAVSTLSLENNTTDNPLDVENKQTAG 415
DB 347 HEMDPLQKPRQRYALRTSPQWLGQIEVIRATKSIEREINSVNDNPLDIVSRNKAHG 406
QY 416 GNGQAAVSSIMKETLALALIGKLNFTQCTELNAAAMRGPSCL-AAEDPSLNYHKG 474
DB 407 GNGQGTPLGVSMKNTLALRALIGKLPFAQFSELVNDFTNGLPSTLTKSRNPDLGKYG 466
QY 475 LDHIAAYASELGHANPVTTFVQPAEMGNQAVNSIALISARTAEANDVLSLLASHLY 534
DB 467 ARIAMASYCELDQYLANPVTSHQSAECHNQDVNSLGLISSKTSSEAVDILKMTSTFLV 526
QY 535 CTIQAVIDLAMELDFKQDFPLPTLLOHLGTLGDVNALALEYKALKRKL-----EQ 588
DB 527 AICQAVDLRLLENLNQTVKNTVSQVAKKVLITG--VNG-ELHPSRFCEKDLKRVDRQ 583
QY 589 TTYDLEPRWHDASFATGTVE-----LLSSPSANVTLLTANAKVASAEKAI 638
DB 584 VYIYADDP-----CATYPLIQKLRQVIVDHALVNGSEKNAVSIH-KIGAFBEEL 635
QY 639 S--LTREVR--NRFWOTPSQAPAHAYISPTRVLYSYREBELVQARGGVFVQOQET 694
DB 636 KAVLPKEVAAARAAYNGTSAIP-NRIKECRSYPLRYREBELG-----ELLTGEKYS 689
QY 695 IGSNVSRIYEALIDGRINHVLKML 719
DB 690 PGEFPKVFYALICEGKIIDPMMECL 714
```

RESULT 6
S48726
phenylalanine ammonia-lyase (EC 4.3.1.5) 3 - parsley
C:Species: Petroselinum crispum (parsley)
C:Date: 18-Oct-1995 #sequence_revision 12-Apr-1996 #text_change 16-Jul-1999
C:Accession: S48726; S56035
R:Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
Eur. J. Biochem. 225, 491-499, 1994

A>Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase i:
A:Reference number: S48725; MUID:95010141; PMID:7925747
A:Accession: S48726
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-718 <APP1>
A:Cross-references: EMBL:X81159
R:Appert, C.
submitted to the EMBL Data Library, August 1994
A:Reference number: S56035
A:Accession: S56035
A:Molecule type: mRNA
A:Residues: 1-58, 'E', 60-718 <APP2>
A:Cross-references: EMBL:X81159; NID:9535007; PIDD:CAA57057.1; PIDD:9535008
C:Genetics:
A:Gene: PAL3
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:204-206/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F:205/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.2%; Score 1017.5; DB 2; Length 718;
Best Local Similarity 36.5%; Pred. No. 1.4e-59;
Matches 269; Conservative 128; Mismatches 275; Indels 65; Gaps 19;

```
QY 14 NGFTGSHAPTKSAGTSLRTRP-----GLDGHAAHQSOLEYQELSDPTDDVVELS 69
DB 5 NGTTNG-HA---NNGNGLDCKKEDPLNMGVAEALITGSHLDVKNMVAEYKRPVYKLG 59
QY 70 GYSLTVRDVVG-AARKGRVQNDDEIRARVDKSVDFLKAQLQNSV--YGVTTGGGSA 126
DB 60 GFTLLISQVAALISARDGSGVKELEBARAGVKASDWMVSMKKTDSYGVTTGGFATS 119
QY 127 DTRTEDANSLOKALIEHQLCGVTPSKSSFSVCGLE---NTLPLEVNGAMVIRVNSLT 183
DB 120 HRTTQGGALQCELRFLNAGI-----FSGGAAGNNTLPHSATRAMMIVRINTLL 170
QY 184 RGHSAVRLVLEALTNFNLNHRITPVLPGSISASGDLSPSYAGATGHPVKVHLA 243
DB 171 QGVSGIRFEILEITIKFLNNITPCLPRGTTASGDLVLSYIAGILGRPSKA---V 227
QY 244 EGTEKIMARAEALSLFGLA--VVLGPKGGLVNGTAVASAMATLSLHDSMLLSQA 301
DB 228 GPTGVTLSPFEAFKLAGVGGFPELOPEGALVNGTAVGSGMAGVVFANILAVLAEV 287
QY 302 LTALVEMVGGQGSFAPFIDVCRPHGQVEVARNIRTLSSGSAVEHEEVEVKVDE 361
DB 288 VSAIYFAEWGKRP-EFTDHLTHLKHHPQCEIAAAMEHILDSGAY-VKAAQGLHEMDPL 345
QY 362 GILRODRYPLRTSPQFLGVLVEDMMAVSTLSLENNTTDNPLDIVENKQTAGNFQAS 421
DB 346 QKPKQDRYALRTSPQWLGQIEVIRATKSIEREINSVNDNPLDIVSRNKAHGKFFQGS 405
QY 422 AVSISMEKTRIALALIGKLNFTQCTELNAAAMRGPSCLA-AAEDPSLNYHKGDLNHA 480
DB 406 PIGVMDNTRIALALIGKLPFAQFSELVNDFTNGLPSTLTKSRNPDLGKYGKAFIYNA 465
QY 481 AAYASELGHANPVTTFVQPAEMGNQAVNSIALISARTAEANDVLSLLASHLYCTLOAV 540
DB 466 SCYSELQFLANPVTSHQSAECHNQDVNSLGLISSKTSSEAVELIKMSTTFVAGCQAI 525
QY 541 DLRAEMLDFKQDFPLPTLLOHLGTLGDVNALALEYKALKRKLQ-----588
DB 526 DLRLLEENLKSTVNTVSQVAKRVLTWG--VNG-ELHPSRFCEKDLRVVDREYIYAYID 582
QY 589 ---TTTYDLEPRWHDASFATGTVEVLLSSPSANVTLLTANAKVASAEKALS--LTRE 643
DB 583 DPCSATYPLMQKRE-----TLVEHALNNGDERLSTISIQKIAAFEDDELKALLPRE 635
QY 644 VRN-RFWOTPSQAPAHAYISPTRVLYSYREBELVQARGGVFVQOQETIGSNVSR 702
DB 636 VETARAALASGNPAIPNRIKECRSYPLRYKREBELG-----EYLTGEKYSRGESEK 690
```

RESULT 7

phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
C:Accession: S66343
R:Fukasawa-Akade, T.; Kung, S.; Watson, J.C.
Plant Mol. Biol. 30, 711-722, 1996
A>Title: Phenylalanine ammonia-lyase gene structure, expression, and evolution in Nicotiana glauca
A:Reference number: S66343; MUID:96194455; PMID:8624404
A:Accession: S66343
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-715 <FUNK>
C:Genetics:
A:Cross-references: EMBL:M84466; NID:G170349; PIDN:AAA4122.1; PII:G170350
A:Gene: PAL1
A:Introns: 133/2
C:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:201-203/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F:202/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.2%; Score 1016.5; DB 2; Length 715;
Best Local Similarity 36.0%; Pred. No. 1.6e+59;
Matches 270; Conservative 130; Mismatches 259; Indels 91; Gaps 19;

DY 13 ANGFTN-GSHAAPKSAAGPTSLRRTPGLDGAHQSOLEIVQELSDPDDVVELSGY 71
 :
DB 3 SNHGHNNGENPELCCKSADPLNWEMAAESLRG-----SHLDVVKAKSEPRKKPVKLGE 57
DY 72 SLTYRDVVGAA---RKGRRVYQNDDELRAVDYSVPFLAKQLNSV--YGTTDFGGSA 126
 :
DB 58 SLTYAQVAALAVRDKSANGVVELSEERRARCVKASSVMWDSNMKGTDSSGYTTFGATGS 117
DY 127 DTRPEDAVSIQKALEIHQICGVTPTSXSFSVGRGLENTPLEVVRGAMYLIRVSLRGH 186
 :
DB 118 HRRIKNGALQKLEIRFLNMGV-----FENGLETHTLPHSATRAAMLVRINTLLQGY 170
DY 187 SAVRLVLVALTNFLNRITPIVPLRGSISASGOLSPSYIAGAITHGPDKVHYLHEGT 246
DB 171 SGRIFFELIEATTKLININITPCLEPRGITTSAGLVPSYAGILLTRPNPSKA-VGNG- 228
DY 247 EKIFAREEALISLFGLBA--VVIKPGREGILVNGTAVASMSMTLSLHDSHMLSLSOALTA 304
DB 229 -ETINAERAFVAVGVNGGFELQKEGIALYNGTAVGSMASMYLPFSNLIAWSEVISA 287
DY 305 LTVAMVQQGQSFAPFIHDVCRPHPGQVEVARNIRTLSSGSFAVEHEEVKVKDEGITL 364
DB 288 IFAEVMNGKP-EFTDHLTHKLKHHPGOIEAAILMEHILDGSSY-VKAQAQKLEMDPIQKP 345
DY 365 RODEYPLRTSQFGLPLVEDMMHAYSTLSLENNTTNPDLIDVENKQTACAGNCQASAS 424
DB 346 KODRIALTSIQMIGPOLIEVIRAATKMIEREINSVNDPILDIVSRNKALHOGNQGTPIG 405
DY 425 ISMKETRIALLIGLNFTQCTELLNAMNRGLPSCU-AAEDPSLYNHGKGLDIHAAYA 483
DB 406 VSMONARIALSIGTLMPAQFSELYINDYNNGLPSNLIASANPSLDGCFKAEIAYMASYC 465
DY 484 SELGHIANPVTTFPQAPMGNOAVNSLALISARKTAEANDVLSTLSLASHLYCTIOAVDLR 543
DB 466 SELQFLAPVNVNHVQSAQHNOADVNSJGLISARKTAEVNDILKLKMSSTYVALCOQADLR 525
DY 544 AMEDPFKQGDPPLPTLLQOHLGTLGVNALALFEKALKLNRELOTTYYDDEPMHDA-- 601
DB 526 HLEENLK-----NA-----VKNIYQVARRLTLMANGELHPARF 560
DY 602 -----FSYATG-----TWELLSSPSANVTLLAVANAKVAS 633

Db	561	CEKELLRIIVREYLFAYADDPCGCNYPFLMOQKLRQVLVDHAMNNGSEKUNVSSIFQIGG	620
Oy	634	AE---KALISTREVRN-RFQWTPSSQAFAHAYLSPTFRLYLSVREELGVQARGVFG	669
Db	621	FEDELKAV-LPKVESAPARALSGNPALFNRIETCKRSYPLFRVRELGT---	674
Oy	690	VOEETIGSNVSRIVEAIKDGRIINHLVVKML	719
Db	675	EKVRSPEGECDKVFATMANGQIIDPMLTECL	704
RESULT 8			
	QJ2265	phenylalanine ammonia-lyase (EC 4.3.1.5) - western balsam poplar x cottonwood	
	C/Species:	Populus trichocarpa x Populus deltoides (western balsam poplar x cottonwood)	
	C/Date:	30-Sep-1993 #sequence_rev150 20-Aug-1994 #text_change 28-May-1999	
	C/Accession:	JQ2265	
	A/Accession:	JQ2265	
	A/Molecule:	type: mRNA	
	A/Residues:	1-715 <SUB>	
	A/Cross-references:	GB:J11747; NID:G169453; PIRN:AAA33805.1; PID:G169454	
	A/Note:	The authors translated the codon AAG for residue 331 as Glu and AAG for residue	
	C/Comment:	The enzyme plays a role in linking primary metabolism to phenylpropanoid meta	
	C/Superfamily:	histidine ammonia-lyase	
	C/Keywords:	ammonia-lyase; carbon-nitrogen lyase	
	F:201-203/Cross-link:	5-imidazolone (Ala-Gly) #status predicted	
	F:202/Modified site:	dehydrotropane (Ser) #status predicted	
	Query Match	28.1%; Score 10.6; DB 2; Length 715;	
	Best Local Similarity	35.8%; Pred. No. 1, 8e-59;	
	Matches	266; Conservative 132; Mismatches 276; Indels 68; Gaps 18;	
Oy	8	LATTIANGFTNGSHAAPTKSAAGPTSAIRTPGLDGHAAHQSOLEIYOELLSPDTDVE	67
Db	1	MEYTKNGYNGSLSEICVQDRPLSN----	55
Oy	68	LSGYSLTRDVG-AAKGRRVYQNDDEIRAVDKSVFLKALQNSV--YGVTTGGG	124
Db	56	LGGETTLIAQVASIGADHTGDVKEYELSESAPVKKASDWMQSMKGTDSYGVTTGGG	115
Oy	125	SACRTEDAVSLKALIEHQLGCVTPSKSFSVGLE--NTLPLEVVGAWVIRNST	182
Db	116	TSRRTRQGGALQDELRFNAGI-----FGNGTECHILPHSATPAALVIRNTL	166
Oy	183	TRGSHAVRLVLEALTNFLNHRITTPVPLRGSISASGDLSELSYAGATGHPDVYVL	242
Db	167	LQVSGIRFEILAEIRLLNNNITPCLPLEGTTTASGDLPLSYIAGLITGRPNRKA--	223
Oy	243	HEGTEKMPAREALISLFGLEA--VLGPKGGLVNGTAVASAMATLSLHSHLSLQ	300
Db	224	TGPFGEVLDAEAFKAGLSGFEPLQPKGLVNGTAVAGSGSLASVLEETNVLAVLS	283
Oy	301	ALATLVEAVVGQGSFAPFIHDVCRPHQGVAVARNIRLLSGSPFAV---HEEYK	356
Db	284	LLSAIFAEVWNGKP-BETDHLTHKLKHHPQIDEAAMIEHILLDSAYKAKXKLHETDPL	342
Oy	357	VKDEGLIRQDRPLRLRSPFLGRLVDDMMHAYSTLSLENNTTTNDPLDVENKQTLHGG	416
Db	343	QKP-----KODRAVLRTPQMLGPQIEIVIRFSTKSIETREINSVNDNPLDIVSRNKAHGG	397
Oy	417	NFQASAVISMEKRLALATLIGKLNFTQCETLNAANRGLPSC-LAAEDPSLNHYHGKL	475
Db	398	NFGQTPPGVMDNVRILASIGKLLFAGPSFLVNDPFPNNLPLNSLRNPSLDYFGKA	457
Oy	476	DIIIAAYASIELGLANVTTPVQPAEMGNQAVNSLALISARRTAEANDVLSLLASGLYC	535
Db	458	EIAAASVCEILOYANVTTHVQSAEQHNDVNSGLISRRKTAEEADVIKIMSTTFLVA	517

QY 536 TLQAVDLRAMELDFKKQFDELLPTLLQOHLGTGLDVNALALEYKALNKRLEQ----- 588
 Db LCQALIDRLHEENLKSAVKNTASQVSRVLTGTANGE---LHPSRFCEKELLKVVDEEV 574
 QY 589 -----TTTYDEPFMHDAFSYATGTVVELSSPSANVLTAVNMAKVASAKAIS- 639
 Db 575 FAYVDDPCSATYPLMQRLQGV-----VDHALENGENENKFTSTV-FQKLEAFEEELKA 627
 QY 640 -LTREVN-RFWQTPSSQAPAHAYLSPRTEVLYSFVEELGVQARGDVFGVQOETIGS 697
 Db 628 LLEPEVASAAYDSGNSA-DNKIXECRSYPLKFRLELGV-----LITGEXQSPGE 682
 QY 698 NVSRIVFAIDGRINHYLVML 719
 Db 683 EFDKFTAMCQGRKIDEMLECL 704

RESULT 9

T46172
 phenylalanine ammonia-lyase (EC 4.3.1.5) 2 [similarity] - Arabidopsis thaliana
 N/Aleternate names: protein T4D2.190
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
 C/Accession: T46172; S52991
 R/Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
 submitted to the Protein Sequence Database, December 1999
 A/Reference number: Z23025
 A/Accession: T46172
 A/Molecule type: DNA
 A/Residues: 1-717 <NMA>
 A/Cross-references: EMBL:AL132958; NID:96434245; PIDN:CA664229.1; PID:g6630746
 A/Experimental source: cultivar Columbia; BAC clone T4D2
 R/Matner, L.A.; Li, G.; Ware, D.; Somssich, I.E.; Davis, K.R.
 Plant Mol. Biol. 27, 327-338, 1995
 A/Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
 A/Reference number: S52990; NID:95195160; PMID:7886622
 A/Accession: S52991
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-205/'H', 207-234,'E', 236-717 <MAN>
 A/Cross-references: EMBL:U33678; NID:9497420; PIDN:AAC18871.1; PID:g497421
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
 C/Genetics:
 A/Gene: PAL2
 A/Map position: 3
 A/Intons: 135/2
 A/Note: T4D2.190
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/203-201/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/204/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.1%; Score 1015.5; DB 2; Length 717;
 Best Local Similarity 36.0%; Pred. No. 1.9e-59;
 Matches 264; Conservative 140; Mismatches 284; Indels 45; Gaps 17;
 QY 5 LDSLATTIANGFTNGSHAAFTSAAGPTSAALRRTPGLGHAHAGOLEIYOELLSPDTPDD 64
 Db 1 MDQIEMLCGGSEKTKVAATTKTLADPLNM-----GLAADQKSHLDEVKKWDEVRFP 55
 QY 65 VVELSGYSLTVADYVGAARKRRVQVNDERARVDKSVFLKAQLQNSV--YGVTTGF 122
 Db 56 VVMIGGETLITIGVAAISITVGGSVKVELAETSRQVKAASDPMVMSNMKGTDSYGVTTGF 115
 QY 123 GGSADRTEDVAVSLQALLENHQLGCVTPTSXSSPFGVGLNLTPLLEVRRAMVIRNSL 182
 Db 116 GATSHRRKTKGTLQCELIRFLNAGT-----FQNTKETCHTLPQSATRAAMVVRVNTL 168
 QY 183 TRGSAVRLVLEALTNPLNHRITPLVPLRGSISASGDLSPSLYAGAITGHPDVKVV 242
 Db 169 LQYSGIRFEILFALITSLNHNISPLPLRGITITSGDLVPLSYLAGLITGRPSKA--- 225
 QY 243 HEGTE-KIMFAEALISFLGLA--VTLGPKEGELGLVNGTAVSASMAITLSLHDSIMLSLS 299

Db 226 -TSPDESILTAAEFERKAGISTGFEDLQPKREGALVNGTAVGASMAVLEFANQVAVLA 284
 QY 300 QALTLATVAMVQGGQSPAPFIHDYCRPHQGVAVARIRTLTGSSFAVHEEVYK 359
 Db 285 EYLSALFAVMSGKR-EFTDHLTRLKHPPQIEAALMEHILDGSSY-MQLAQVHEMD 342
 QY 360 DEGITLRQDEYPLRTSPDFGLPVEDMMAVSTLSLNNTTTDDNLDLVENKQTHAGNFQ 419
 Db 343 PLQKRPQDRAVLRTPSPWLGPIEVRQATSIEREINSVNDNLLIVSRKAIHGNFQ 402
 QY 420 ASAVSISNEKTRPLALILGKLNFTQCTLLMAANRGIPCL-AAEDPSLNYHGKJDIH 478
 Db 403 GTPIGVSMNTRLAALIGKLMFAQFSELVDFYNNGLPSNLTASSNPSSLDYGFKAELA 462
 QY 479 IAAVASELGHANPVTTFVQAPEMNGQAVNSLALISARTAEANDVLSLASHLYCTLQ 538
 Db 463 MASYCELOYLANPVTSHVQSEHQHODVNSLGLISSKTSIBAVDILKMSSTTLVIGICQ 522
 QY 539 AYDLRAMELDFKKQFDELLPTLLQOHLGTGLDVNALALEYKALNKRLE-----EQTTY 592
 Db 523 AYDLRLHEENLQTVKNTVSGVAKKVLTTGNGE---LHPSRFCEKDLLKVVDBEQVFTY 579
 QY 593 DLEP-RHMDASVATGVVE-LSSSPSANVLTAVNMAKVASAKAIS--LTREVRN 647
 Db 580 VDDPCSATYPLMQRLQGVTVDAHSNGETRNAYTSI-FQKIGAFEEELKAVLPKEVEAA 638
 QY 648 FWTQTPSSQAPAHAYLSP-RTVLYSFVEELGVQARGDVFGVQOETIGSNVRIVEAI 706
 Db 639 RAAYNGGARPIRNRKERSYPLRFVAREELGTLLIGEXV-----SPSEBFDKXVTAM 693
 QY 707 KQGRINHYLVML 719
 Db 694 CEKRLIDPLMDCL 706

RESULT 10

S22991
 phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - soybean
 C/Species: Glycine max (soybean)
 C/Date: 06-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
 C/Accession: S22991
 R/Frank, R.L.; Vodkin, L.O.
 DNA Seq. 1, 335-346, 1991
 A/Title: Sequence and structure of a phenylalanine ammonia-lyase gene from Glycine max
 A/Reference number: S22991; NID:92190550; PMID:1799682
 A/Accession: S22991
 A/Molecule type: DNA
 A/Residues: 1-713 <FRA>
 A/Cross-references: EMBL:X52953; NID:918376; PIDN:CA47129.1; PID:g18377
 C/Genetics:
 A/Gene: PAL1
 A/Intons: 131/2
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/199-201/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/200/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.0%; Score 1011; DB 2; Length 713;
 Best Local Similarity 36.9%; Pred. No. 3.8e-59;
 Matches 270; Conservative 120; Mismatches 282; Indels 60; Gaps 19;
 QY 14 NGFTGSHAAFTSAAGPTSAALRRTPGIDGAAHAGOLEIYOELLSPDTPDVELSGYSTL 73
 Db 5 NCHQNGSTCLST-AKGNPDPLNM--GAALFAMKSHLDEVKKMAVEYKRPVRLIGETTL 60
 QY 74 TVRDVGAARKRRVQVNDERARVDKSVFLKAQLQNSV--YGVTTGFGGSADTRTE 131
 Db 61 TIAQVAAYAGHDGVAVELSESAREGVKASSEWMSNMNNGTDSYGVTTGFGATSHRRK 120
 QY 132 DAVSLQKALLIHQOLGCVTPTSXSSPFGVGLNLTPLLEVRRAMVIRNSLTRGSAVRL 191
 Db 121 QGGALQKELIRFLNAGT-----FQNGTSSSHLTPITRAAMVIRINILLOQYSGIRP 173


```

OY 9 ATTLANGPTNGSHAAPTSAAGPTSLARRTP-----GLDGHAAHQSOLEIVQELLSPTD 64
   |||||
Db 7 ATT--NGHVNGN-----GMDPCWKTEDPLVWGIAAEAMTGHSLDDEVKQVAVEYRK 55
   |||||
OY 65 VVELSGYSLTVRDVVG--AARKGRVRYQNDDEIRAVDKSVDFLKAQLQNSV--YGVTTG 121
   |||||
Db 56 VVKLGSETLTISQVAAIASARDGSGVTVEISEAARAGKASSDVMWDMKKGTDSYGVTTG 115
   |||||
OY 122 FGGSADTRTBDAVSIQKALIEHQLOGVPTXSXSFVGRGLENTLPLEVVRGAMVIRVNS 181
   |||||
Db 116 FGATSHRRTKQGQALQKELIRPLNAGI-----FGNGSDVTLPHSATRAAMLVRIINT 166
   |||||
OY 182 LTRGSAVRLVLEALTNFLNHRITPIVLRGSIASGDLSPSYIAGATIGHPDVAVH 241
   |||||
Db 167 LLQGYSGIRFEILEAIFKFNQITPCLPLRGITIASGDLVPLSYIAGLITGRPNKA-- 224
   |||||
OY 242 LHEGTEKTFAREALISFLGEA--VVLGPKEGIGLVNGTAVGASMATLSLHSHMLSLIS 299
   |||||
Db 225 -VGPTGYILSPEEAFKLAGEGPFELQPKKEGALVNGTAVGASMAWMLPEANILAVLA 283
   |||||
OY 300 QALTALVEAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEEYKXD 359
   |||||
Db 284 EVMFAIPAEVWQGR--EFTDHLTKLKHHPGQIEAALIMEHILDGSAV--VKAQQLHEMD 341
   |||||
OY 360 DEGILRODRYPLRTSPQFLGLVEDMMAVSTLSLENNTTDNPILLDVENKQTAHGNFQ 419
   |||||
Db 342 PLQPKQDRYALRTSPQFLGLVEIRRSSTKTEREINSVNDNPLIDVSRNKALIGSNFQ 401
   |||||
OY 420 ASAVISMEKTRIALALIGLNFQCTELLNAAMNRLPSCLA--AEDPSLNYHGKLDIH 478
   |||||
Db 402 GTPGVGMDTRLAIAIGKLMFAQFSELVNDPYNNGLSGNGNPLDYGFGAEIA 461
   |||||
OY 479 IAAVASLGLHNPVTTFVQPAEMGNQAVNSIALISARTRANDVLSLLASHLYCTLO 538
   |||||
Db 462 MASCSSELQPLANVTNHVQSAEQHODVNSIGLISRSKTSSEVLELKLMSGTFLVLGCLQ 521
   |||||
OY 539 AVDIRAMELDFKQDFPLPTLLOQHIGTGLDVNALALEVKALKRLEQ----- 588
   |||||
Db 522 AIDRHLIEENLKSVYKTVSSVAKRYLWNG--VNG--ELHPSRFCKDLIRVVDREYFAY 578
   |||||
OY 589 -----TTTYLLEPRMEDAFSAYATGYVELISSPSANVTLLTAVNMAKVASAKAIS--LT 641
   |||||
Db 579 IDDECSATYPLMQLRQ-----TLVEHALKNGDNERNLSTISPOKIAATEFDEKALLP 631
   |||||
OY 642 REYVN--RFWQTPSSQAPAHAYLSPTRVLYSPFRELEIGVQARRGDPVGVQGETIGSVVS 700
   |||||
Db 632 KEVESAAALAESGNPAIPNIECRSYPYKFKRKELGT-----EYLNGEKVTSFGEEFE 686
   |||||
OY 701 RIYEALIDGRINHVLVKML 719
   |||||
Db 687 KVFIAKSGEIIIDPLECL 705
   |||||

```

RESULT 13

S25303
phenylalanine ammonia-lyase (EC 4.3.1.5) - garden pea
C/Species: Pisum sativum (garden pea)
C/Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 20-Jun-2000
C/Accession: S25303
R/Kawata, S.; Yamada, T.; Tanaka, Y.; Stripaserisak, P.; Kato, H.; Ichinose, Y.; Kato, Plant Mol. Biol. 20, 167-170, 1992
A/Title: Molecular cloning of phenylalanine ammonia-lyase cDNA from Pisum sativum.
A/Reference number: S25303; MID:92385763; PMID:1515609
A/Molecule type: mRNA
A/Residues: 1-723 <KAW>
A/Cross-references: EMBL:DJ0001; NID:g217979; PIDN:BA00865.1; PID:g217980
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F/209-211/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F/210/Modified site: dehydroalanine (Ser) #status predicted

```

Query Match 27.9%; Score 1005.5; DB 2; Length 723;
Best Local Similarity 35.8%; Pred. No. 9e-59;
Matches 270; Conservative 135; Mismatches 258; Indels 91; Gaps 23;

OY 9 ATTLANGF-----TNGSHAAPTSAAGPTSLARRTPGDDGHAAHQSOLEIVQELLSPTD 63
   |||||
Db 7 AITKNGYISFCVTYAKNNMKVNSADPLN-----GYAALAMGSHLDEVKRYAVEYRK 61
   |||||
OY 64 DVEISGYSLTVRDVVGAAARKGRVRYQNDDEIRAVDKSVDFLKAQLQNSV--YGVTTG 121
   |||||
Db 62 PVVRLGSETLTISQVAAIAADHGVKVELSEASAGVASSDVMWDMKKGTDSYGVTTG 121
   |||||
OY 122 FGGSADTRTBDAVSIQKALIEHQLOGVPTXSXSFVGRGLENTLPLEVVRGAMVIRVNS 181
   |||||
Db 122 FGATSHRRTKQGQALQKELIRPLNAGI-----FGNGTSSHTLPHSATRAAMLVRIINT 174
   |||||
OY 182 LTRGSAVRLVLEALTNFLNHRITPIVLRGSIASGDLSPSYIAGATIGHPDVAVH 241
   |||||
Db 175 LLQGYSGIRFEILEAIFKFNQITPCLPLRGITIASGDLVPLSYIAGLITGRPNKA-- 232
   |||||
OY 242 LHEGTE--KIMFARAL-----SLFGEAVVLGPKEGIGLVNGTAVGASMATLSLHSH 293
   |||||
Db 233 ---GTSGEILINKKEAFQSAINDGFEELQ-----PKELALVNGTAVGSGLASIVLEAN 284
   |||||
OY 294 MLSLSQALTALTVAMVGQGSFAPFIHDVCRPHPGQVEVARNIRTLSSGSPFAVEHEE 353
   |||||
Db 285 ILAVSEVLSAIPAEVWQGR--EFTDHLTKLKHHPGQIEAALIMEHILDGSAV--VKAQ 342
   |||||
OY 354 EVKVKDDBSILRODRYPLRTSPQFLGLVEDMMAVSTLSLENNTTDNPILLDVENKQTA 413
   |||||
Db 343 KLHEMDPLQPKQODVYALRTSPQFLGLVEIRRSSTKTEREINSVNDNPLIDVSRNKAL 402
   |||||
OY 414 HGSNPOASVSIKMEKTRIALALIGLNFQCTELLNAAMNRLPSCL--AEDPSLNYHG 472
   |||||
Db 403 HGSNPOGTPIGVSMONTRIALALISGKLFAQFSELVNDPYNNGLSGNGNPLDYGFGAE 462
   |||||
OY 473 KGLDIIAAVASLGLHNPVTTFVQPAEMGNQAVNSIALISARTRANDVLSLLASH 532
   |||||
Db 463 KSEIMASGYSELQYLANPVTTHVQSAEQHODVNSIGLISRSKTSSEVLELKLMSGTFL 522
   |||||
OY 533 LYCTIQAVDIRAMELDFKQDFPLPTLLOQHIGTGLDVNALALEVKALKRLEQTTY 592
   |||||
Db 523 LIALQCAVDRLIEENLKSVKNTVSVAKRYLITG--VNG--ELHPSRFCKDLIRV-- 577
   |||||
OY 593 DLEPRWDAFSY-----ATGTAVV-----LLSSPSANVTLLTAVNMAKVASAEK 636
   |||||
Db 578 ---DNEH--VFAYIDPCGATYPLMQLRQVLVDHALVNGSEKRLNLSIFQ--KIAFFED 631
   |||||
OY 637 AIS--LTREVRNRFWQTPSSQAPAHAYLSP-----RTRVLYSVREELGVQARRGD 685
   |||||
Db 632 ELKTLPLKEV-----ESTRAAVESGNPVPNPKINGCRSYPYRFEQELGTGLTJGE 683
   |||||
OY 686 VFGVQOERTIGSNVRIEAIKDGRIHVLVKML 719
   |||||
Db 684 KVI-----SPGSECDKLTALQCGKTIIDPLQCL 712
   |||||

```

RESULT 14

T03663
phenylalanine ammonia-lyase (EC 4.3.1.5) - common tobacco
C/Species: Nicotiana tabacum (common tobacco)
C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Jul-1999
R/Pellegrini, L.; Rohlfitech, O.; Fritig, B.; Legend, M. Plant Physiol. 106, 877-886, 1994
A/Title: Phenylalanine ammonia-lyase in tobacco. Molecular cloning and gene expression
A/Reference number: Z14995; MID:95125127; PMID:7824656
A/Accession: T03663
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-712 <PEL>
A/Cross-references: EMBL:X78269; NID:9633596; PIDN:CAA55075.1; PID:9633597
A/Experimental source: cultivar Samsun NN, tissue-type leaves after infection by tobac

C:Genetics:
 A:Gene:PAL
 A:Function:
 A:Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid
 A:Note: strongly induced during the hypersensitive reaction to TMV or to a fungal elicitor
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 F:198-200/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:199/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.8%; Score 1005; DB 2; Length 712;

Best Local Similarity 37.4%; Pred. No. 9.5e-59;
 Matches 260; Conservative 131; Mismatches 253; Indels 52; Gaps 17;

49 SOLEIYQELISPTDVEVLSGLTVRDVVGARK--GRVRYQNDDEIRARVDKVD 106
 33 SHLDEYKVAERKRVVVLGGELITVACVAALAKDKAKYKVELSGARAGYKASDW 92
 107 LKAOLQNSV--YGVTTGFGSADRTEDAVSLQKALIEHOLCGVTPTSXSSFSVGRL 163
 93 VMDSMKGTDSYGVTTGFGATSHRRTKNGALQKELIFLNAGV-----FNGTGES 143
 164 -NTLPLEVRGAMVIRVNSLTGRGSAVHVLVLEALTNLNRITPIVPLRGSISAGDLS 222
 144 CHTLPGSGTRAMLVIRINTLLQGYSGIRFELIEATLKLNHVPLCLPRGTTASGDLV 203
 223 PLSYIAGATGHPDVVHVLHEGTRKIMFARPAISLFGLEA--VVLGPKGELGVNGTAV 280
 204 PLSYIAGLTGRPNKA--IGPNG--ETLNAEAFVAGVNSGFEPLQKEDALVNGTAV 260
 281 SASMATLSHDSHMTLSLQALTALTVEAMVQGGSPAFPHDVCRRHPGQEVANIR 340
 261 GSGIAMVFLFANITLAVFSEVLSAIFAEVWNGKP--EFTDHLTHKLKHHPQOIEAAATHE 319
 341 LSGSFAVHEHEEVKVKDDEGILRODRPLRTSPQFGLVEDMMAHYSTLSLNNTTT 400
 320 ILDGSSY-VKAPQKLEHTPLQPKQDYALTSQMLGPOLEVRKATKMIEREINSVN 378
 401 DNELDVENKQTAAGNFPQASAVSISMEXTRLALIGKLNFTQCTELNNAANGLPSC 460
 379 DNELDVSNRKALHGNFGQTPIGVSMVARLALASIGKLMFAQSELVNDYNNGLPSN 438
 461 L-AAEDPSLTHGKGLDITIAAYASGLHLANPYTPVQPAEMGNQAVNSLALISART 519
 439 LTRGRNPSLDYGRKSEIMASYSCELOPLANPVNHHVQSAEQNHQDVSIGLISART 498
 520 EANDVLSLASLHLYCTLOAVDLRAVELDFKQFPLLPFTLLOCHLGTGLVNALALEVK 579
 499 EAVDILKMSSTLVLCQALIDLRHLEENLRANVNTVSQVAKRTLTMGANGE---LHPS 555
 580 KALNKELEQTTVDLEPRHDAFSYATGVELL-----SSPSANVTTLAVN 627
 556 RFCEKDLRVVDREYFRYADACSNYPMLQKLOVLDHALENGEMKXANSIFQ-- 613
 628 AMVVASAE--KAISLTREVRN-RFMQTPSSQAPAHAYLSPRTRLVSFYVEBELGVQARR 683
 614 --KLMAEGELKAV-LPKEVESARISLENGENPALARIKESYFLYRFVREELG----- 665
 684 GDVAVQOCTTIGSNVRIYEAIKDGRIHVLVKML 719
 666 AELLTGKXVSPGEBQKVFTAMCNGQIIDSLECL 701

RESULT 15

S48725
 phenylalanine ammonia-lyase (EC 4.3.1.5) 2 - parsley
 C:Species: Petroselinum crispum (parsley)
 C>Date: 28-Oct-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
 C:Accession: S48725
 R:Appert, C.; Logemann, E.; Hahlbrock, K.; Schmid, J.; Amrhein, N.
 Eur. J. Biochem. 225, 491-499, 1994
 A:Title: Structural and catalytic properties of the four phenylalanine ammonia-lyase iso
 A:Reference number: S48725; PMID:95010141; PMID:7925471

A:Accession: S48725
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-716 <APP>
 A:Cross-References: EMBL:X81158; NID:9534892; PIDD:CA547056.1; PIDD:9534893
 C:Genetics:
 A:Gene: PAL2
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 F:202-204/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:203/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.8%; Score 1005; DB 2; Length 716;

Best Local Similarity 35.8%; Pred. No. 9.6e-59;
 Matches 264; Conservative 133; Mismatches 271; Indels 70; Gaps 19;

12 LKNGFTNSHAPTSAGPTSAALRTP---GLDGAHQSOLEIYQELISPTDVEV 67
 8 ITNGVHNGN-----GMDFCMKTEDPLYWGLAEBMNGSHLDEYKVAERKRVVK 58
 68 LSGSLTVRADVVG--AARKGRVRYQNDDEIRARVDKSYDFLKAOLQNSV--YGVTTGFGG 124
 59 LGGELITISQVVAISARQSGVTVLSEARAGYKASDWMDSMNKGTDSYGVTTGFGA 116
 125 SADITTEDAVSLQKALIEHOLCGVTPTSXSSFSVGRGIENTLPLEVVRGAMVIRVNSLTR 184
 119 TSHRRTKQKQALQKELIFLNAGI-----FNGSDNTLPPSARPAALVIRINTLLQ 169
 185 GSAVALVLEALTNLNLNRITPIVPLRGSISAGDLSPLSYIAGATGHPDVVHVLHE 244
 170 GYSGIRFELIEATFPLNQNTITPCLPRGTTASGDLVPLSYIAGLTGRPNKA---VG 226
 245 GTEKIMFARPAISLFGLEA--VVLGPKGELGVNGTAVSASMATLSHDSHMTLSLQAL 302
 227 PTVGILLSPEAKYKLAGVGGFEPLQPKKGLALVNGTAVGSAAMVFLFANITLAVLEVM 286
 303 TALYTEAMVQGGSPAFPHDVCRRHPGQEVANIRITLSGSSFAVHEHEEVKVKDDEG 362
 287 SAIFAEVWNGKP--EFTDHLTHKLKHHPQOIEAAATHEITLDGSAV--VKAQKLEHMDPLQ 344
 363 IIRQDRYPLRTSPQFGLVEDMMAHYSTLSLNNTTTNDNELDVENKQTAAGNFPQASA 422
 345 KPQKQDYALTSQMLGPOLEVRKATKMIEREINSVNDNELDVSNRKALHGNFGQTP 404
 423 VQISMEKTRFLALIGKLNFTQCTELNNAANGLPSCLA--AEDPSLTHGKGLDITIAA 481
 405 IGMNDMTFLAIAAIQKLMFAQFSELVNDFYNNGLPSNLGGRNPSLDYGRKGAETAVAS 464
 482 YASELGHLANPYTPVQPAEMGNQAVNSLALISARTAEANDVLSLASLHLYCTLOAVD 541
 465 YGSELOFLANPVNHHVQSAEQNHQDVSIGLISARTSEAVBEIKMGSTTVLVCQALD 524
 542 LRAMELDFKQFPLLPFTLLOCHLGTGLVNALALEVKKALNKELEQ----- 588
 525 LRLEENLKSATKNTVSSVAKRVLTMG--VNG-ELHPSRFCEKDLRVVDREYIFAYIDD 581
 589 --TTVDLEPRHDAFSYATGVVE--LSSSPSANVTTLVNMAMKVASAKAIS--LTR 642
 582 PGATVPLMLQKLRQ-----TLVEHALKNQDNERNKNTSIFQ--KATTFDEELKALIPK 632
 643 EVAN-RFMQTPSSQAPAHAYLSPRTRLVSFYVEBELGVQARRGDVAVQOCTTIGSNVR 701
 633 EYSSAALAESGNPALPNRIECSRYPYKKFVKELGI-----EVLTGKXTSPGEBEDX 687
 702 IYEAIXDGRINHVLVVKML 719
 688 VFLAMSGEIIIDPLECL 705

Search completed: September 9, 2004, 10:08:45
 Job time : 32.8797 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:50:23; Search time 16.4315 Seconds
(without alignments)

2281.620 Million cell updates/sec

Title: US-09-939-408a-13
Perfect score: 3610
Sequence: 1 MAPSDSLATLNGFTNGS.....RIYAIKGRINHYVXLA 720

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	261.4	72.4	716	1	PALY_RHOTO
2	2499.5	69.2	713	1	PALY_RHOTO
3	1346.5	37.3	740	1	PALY_AMANT
4	1247.5	34.6	724	1	PALY_USITMA
5	1035.5	28.7	725	1	PALY_ARATH
6	1024	28.4	717	1	PALY_PRUV
7	1020	28.3	707	1	PALY_ARATH
8	1019	28.2	725	1	PALY_TRIST
9	1017.5	28.2	718	1	PALY_PETCR
10	1016.5	28.2	715	1	PALY_TOBAC
11	1016	28.1	715	1	PALY_PETCR
12	1015.5	28.1	717	1	PALY_ARATH
13	1013.5	28.1	713	1	PALY_DIGLA
14	1012	28.0	714	1	PALY_CAMST
15	1011	28.0	713	1	PALY_SOYAN
16	1009	28.0	725	1	PALY_MEDSA
17	1006	27.9	716	1	PALY_PETCR
18	1005.5	27.9	723	1	PALY_PEA
19	1005	27.8	712	1	PALY_TOBAC
20	1005	27.8	716	1	PALY_PETCR
21	1004	27.8	715	1	PALY_STYHU
22	1003.5	27.8	712	1	PALY_PHAVU
23	998	27.6	701	1	PALY_ORYSA
24	998	27.6	705	1	PALY_LITTE
25	997.5	27.6	721	1	PALY_LYCES
26	997	27.6	720	1	PALY_SOLTU
27	995.5	27.6	724	1	PALY_LYCES
28	995.5	27.6	724	1	PALY_PEA
29	995	27.6	708	1	PALY_LITTE
30	993.5	27.5	710	1	PALY_TPOBA
31	993	27.5	712	1	PALY_TOBAC
32	991	27.5	718	1	PALY_CICAR
33	987.5	27.4	667	1	PALY_HELAN

34	983	27.2	708	1	PALY_DAVCA	O23865 daucus caro
35	982.5	27.2	710	1	PAL2_ORYSA	P53443 oryza sativ
36	980	27.1	710	1	PAL1_RUBID	O98568 rubus iden
37	979.5	27.1	722	1	PALY_CITUL	O42667 citrus limo
38	973.5	27.0	700	1	PALY_WHEAT	O43210 triticum ae
39	973.5	27.0	754	1	PALY_PINTA	P53777 pinus taeda
40	966	26.8	590	1	PAL2_SOLTU	P31426 solanum tub
41	964.5	26.7	695	1	PAL3_ARATH	P45725 arabidopsis
42	954.5	26.4	707	1	PAL1_TPOBA	P14166 ipomoea bat
43	953.5	26.4	710	1	PAL2_POPKI	O43052 populus kit
44	947	26.2	682	1	PAL1_POPKI	P45731 populus kit
45	946.5	26.2	710	1	PAL3_PHAVU	P19143 phaseolus v

ALIGNMENTS

```

RESULT 1
ID PALY_RHOTO STANDARD; PRT; 716 AA.
AC P11544;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
PAL.
OS Rhodospiridium toruloides (Yeast) (Rhodotryula gracilis).
OC Eukaryota; Fungi; Basidiomycota; Uredinomycetes;
OC Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
CX NCBI_Taxid=5286;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 0559;
RX MEDLINE=88112870; PubMed=2828184;
RA Anson J.G., Gilbert H.J., Oram J.D., Minton N.P.;
RT "Complete nucleotide sequence of the Rhodospiridium toruloides gene
RT coding for phenylalanine ammonia-lyase."
RL Gene 58:189-199(1987).
RL -!- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -!- PATHWAY: Phenylpropanoid biosynthesis, first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -!- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC EMBL: X51513; CAA5886.1; -
CC EMBL: M18261; AAA33883.1; -
CC EMBL: X12702; CAA31209.1; -
CC PIR: A29607; A29607.
CC PIR: A56628; A56628.
CC HSSP: P21310; 1B87.
CC InterPro: IPR008948; L-Asparase-like.
CC InterPro: IPR001106; Phe/His NH3lyase.
CC InterPro: IPR005922; Phe_am_lyase.

```


CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC between the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	
DR	EMBL; AJ010143; CAA09013.1; -.
DR	HSSP; P21310; 1B8F.
DR	InterPro; IPR008948; L-Asparase-like.
DR	InterPro; IPR001106; Phe/His NH3lyase.
DR	InterPro; IPR005922; Phe_am_Lyase.
DR	Pfam; PF00221; PAL; 1.
DR	TIGRfam; TIGR01226; phe am lyase; 1.
DR	PROSITE; ps00488; PAL_HISTIDASE; 1.
KW	Lysase; Phenylpropanoid metabolism.
FT	CROSSLINK 224 226 5-indazoleinone (Ser-Gly)
FT	(By similarity).
FT	MOD_RES 225 225 DHA (2,3-DIHYDROALANINE) (BY SIMILARITY).
SC	SEQUENCE 740 AA; 80156 MW; B05CFFDP887BCBE CRC64;
QY	Query Match 37.3%; Score 1346.5; DB 1; Length 740; Best Local Similarity 43.0%; Pred. No. 2e-82; Matches 317; Conservative 131; Mismatches 238; Indels 51; Gaps 15;
DG	10 TTLANGF--TNGSHAAPFKSAGPTSLARTPGDDHAAHQSG-LEIVQELSDPTDVV 25 TPVPNGFATALSRASTHT--KTSAL-----SOFLEYVELGGYKKGRAI 68
DG	67 ELSGYSLTVRDVGGAARKGRRRVQNDDEIRARVDKSVDFL--KAQLONGSYGVTTFGG 124 69 KYDGQGLSIAAVAARAAYNAVALDESPLVKERYRSQALAIANKVSTGASVYGSLTFGG 128
DG	125 SADPTEDAVSLQKALIHQCCTPTPSXSSFV--GRGLENT-LLEVVRAMVIRNVS 161 129 SADRTRDKMFLGPALLQHNVGILLPTSEPLDVLPLQDANNISMEAWIKAILIRMS 188
DG	182 LARGSHAVLVLEALTLPFNHRITTPVPRGSISASGDSPSYAGAITHPPDYKVAV 241 189 LRHGSGIRMELIEKRRELLAANYIPVPRGSISSGDLSPSYAGTIIGNPSIKXY- 247
DG	242 LH---EGTEKMFAREAISLFGEAVVLOPKRGCLVNGTAVSAMATLSHDSMTSL 297 248 -HGFSKSGRQIGSKSDVALAHNIPEPLESKEPGLINGTAFSAVAALANEALIELVL 306
DG	298 LSGALLTALVEVMQQQSFPAFHDVCRHPCQGVETAKRIIFTLGSSFAVHEEEVAV 357 307 LAQCVTAGTBMLITRASHPFLFIHARARHPQOVEAEINWMILDGSLAQDEHEVRL 366
DG	358 KODEGILRDARYPLTSPQTGPLVEDMDMAVSTLSIENN--TTDNPLLDEVNKOTAHG 415 367 EDKYTLQRDDRPLATSPQFGPIEDIIISAFQTVQECNYLPAIDNPILDGTGSHHG 426
DG	416 GNFGASAYSISNEKRLIALIGLNTTCQTELLMANMRGDLPSCLAADPSINVHGXL 475 427 GNFOAAAVYNAREKRLALHHVGLTLFSQSTELVPNMKGDLPPSYAALDPSINTHAKCL 486
DG	476 DHIAAYASELHLNMPVTTFVQAENGNOAVNSALISARRFAEANDVLSLLASHLYC 535 487 DIATAAYAVE----ATPGETHIQAEHNNOAVNSALISARRITSTLEVLTSLIASYLEI 542
DG	536 TLQAVDRAMELDFFKQDPLLPILLQOCHGTGI--DVNALALEYKXALKNRLEDTTY 592 543 LCALDRLALOEFPLGDIIRBELRSSFSFLSSFEQEKLIQONTSAFEDHLDITTM 602
DG	593 DLERRHDFSATGTVELLSSS-----PSANTVLAVAVAAAEKALISTREVRN 647 603 DNTRKITTAAASSVVLIQFTDGCASVPSGSCDLLSVSSPOSSVATSSSVLMIDLKE 662
DG	648 FWOGPSQAPAHAVLSPRTVLVYSFVVEELGVQARRDVF-----VGVOQETIGSNVR 701

Db 663 YIFEDRGPTASOYIG-KTRPVYQPIRTTIGVAKHSEMYNKFYNGLGVEDVTIGONISR 721

QY 702 IYEAIXDGRINHVLVVM 718

Db 722 IYESIRDKMQSITVSL 738

RESULT 4

PALY_USTMA STANDARD; PRT; 724 AA.

ID PALY_USTMA STANDARD; PRT; 724 AA.

AC Q96V77;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Phenylalanine ammonia-lyase (EC 4.3.1.5).

GN PALI.

OS Ustilago maydis (Smut fungus).

OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;

OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.

OX NCBI_TaxID=5270;

RN [1]

RP SEQUENCE FROM N.A.

RA Kim S.-H., Macdonald K., Virmani D., Wake K., Kronstad J.W.,

RA Ellis B.E.;

RT "Cloning and disruption of a phenylalanine ammonia-lyase gene from

RT Ustilago maydis."

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Catalyzes the nonoxidative deamination of L-

CC phenylalanine to form trans-cinnamic acid and a free ammonium ion.

CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).

CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),

CC which is formed autocatalytically by cyclization and dehydration

CC of residues Ala-Ser-Gly (by similarity).

CC -1- SIMILARITY: Belongs to the PAL / histidase family.

CC

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; AF306551; AAC09388.1; -.

DR InterPro: IPR008948; L-Asparaginase-like.

DR InterPro: IPR001106; Phe/His_NH1lyase.

DR InterPro: IPR005922; Phe_am_lyase.

DR Pfam; PF00221; PAL; 1.

DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.

DR PROSITE; PS00488; PAL_HISTIDASE; 1.

KM Lyase; Phenylpropanoid metabolism.

FT CROSSLINK 205 207

FT

FT MOD_RES 206 206

FT

FT MOD_RES 206 206

FT

SEQUENCE 724 AA; 79277 MW; 02A73C97C34CD1AB CRC64;

Query Match 34.6%; Score 1247.5; DB 1; Length 724;

Best Local Similarity 40.8%; Pred. No. 8, 1e-76;

Matches 302; Conservative 130; Mismatches 243; Indels 65; Gaps 18;

QY 23 APTKSAAGPTSAIRTPGLDGHAAHOSQLEIVQELISDPTD-----DVEELSGY 71

Db 2 APTADVLPVEASTR-PG-----LLVQPSDTLKAASFRTQGVVIDGY 44

QY 72 SLTVRDVYGAARKGRVAVONDDERAVYKSVDFELXQLO--NSVGVTTGPGGSADTR 129

Db 45 NLKIQIGIVASARYGHVTRPPASMETKRIIDSVSLIALDGGESIVGINTGFGSADSR 104

QY 130 TEDAVSLOKALIHQOLGVTPTSKSSFSVGR-----GLENT-----LPLEVVRGAMV 176

Db 105 TANTRALQIALLOMOQCVLPVP-STFPTGEBBSAPFLPLPTDRESSLIMEAVRGAIV 163

QY 177 IRVNSLITRGSAVRLVLEALTN-FLNHRITPIVPLRGSSISASGDLSPSYIAGATQHP 235

Db 164 VRLSLMRHSGVGRVIEDVKQKPLQNNVTPVYVRSISASGDLSPSYIAGALAGOR 223

QY 236 DVKRVYTHGTEKI-MFARBAISLFGLEAVYVLRGEGIGLVNAGVASNAVTLSDHSM 294

Db 224 GYICFVTDGRGQVAVKTADACRMKRIIPVQYPEREALGLLNGTAFASVAGLATYAEH 283

QY 295 LSLISQALFALTVEAMVWGQGSFAPFIDVCRPHPGQYVARNRTLSGSSFA--VEHE 352

Db 284 LMSLTUTLTAVAVELKIGDASFAPFIEIRAPHPGQKSKAKFIRAHSSGRALAEHENE 343

QY 353 BEYKRVKDEGILRQDRYPLRTSPQFLGLVEDVWGHANSTLSLENNVTTDPLDV--ENK 410

Db 344 KVLVPSIEDNGTLRQDRYPLQRTASQWVGGLDINAKSVDFEINSTDDPMIDPYDGDG 403

QY 411 CTHAGGNFQASVSI-SMEKTRIALALIGKLNFTQCTELINAMRGIPSCIAA-EDPSLN 469

Db 404 RIHGSNFPQAMNTAVEXIKRIALCAKGMKTFQCMTELINPMRGLPAULASTPLDSLIN 463

QY 470 YHGKGLDIIHAAVASELGHANPVTTFVQPAEKGNQAVNSIALISABRTAEANDVISLL 529

Db 464 FPAKGINIALASVTSELMFELGNPVSTHVQSAEMNQAFNSIALISGRQLQATIECLSMIQ 523

QY 530 ASHLVYCTQAVYDRLAMELDFKQPPILLPTLQCHLGTGD--VNALAEVYKANKRL 586

Db 524 AMSVLICQALDIRLQKVABQPLETILASHFEGMDEYKQCEIQAQVLKMSKRL 583

QY 587 EOTTVYDLEPRMHDASFATGTVVELSSPSANVT--LTPVNAKVASAEKAIISLTREV 644

Db 584 DETSSGLDRDELVEYVQDASSVLYRYSLELPGSGGAPDLINIVYMRATGVADKEIKRYV 643

QY 645 RNRFPWTPSSQAPANA-YLSPETRYLVSFYREELCYQARRD--VVG--VQCEITGSN 658

Db 644 TTEFLDNPYA--CHASHLIGKTRAVYFKTLGVPMGKXENLNEKGFEDQNTGGY 700

QY 699 VSRIEATXKGRINHVLVVM 718

Db 701 VSVIYASIRDSGLYMLSEL 720

RESULT 5

PALI_ARATH

ID PALI_ARATH STANDARD; PRT; 725 AA.

AC P355T0; Q9ZQD6;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).

GN PALI OR ATG37040 OR TLH8.22.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Arabidopsi: Vitididplanteae: Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=cv. Landsberg erecta;

RX MEDLINE=95195160; PubMed=788622;

RA Wanner L.A., Li G., Ware D., Somsich I.E., Davis K.R.,

RT "The phenylalanine ammonia-lyase gene family in Arabidopsis

RT thaliana."

RL Plant Mol. Biol. 27:327-338 (1995).

RN [2]

RP SEQUENCE FROM N.A.

RA STRAIN=cv. Columbia;

RX MEDLINE=20083487; PubMed=10617197;

RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,

RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,

RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,

RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umeyam L.,

RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Mierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
RT *thaliana*.";
RL Nature 402:761-768(1999).
RN (3)
RN SEQUENCE OF 1-240 FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=93005677; PubMed=2152131;
RA Ohl S., Hedrick S.A., Chory J., Lamb C.J.;
RT "functional properties of a phenylalanine ammonia-lyase promoter from
RT *Arabidopsis*.";
RL Plant Cell 2:837-848(1990).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: J33677; AAC1870.1; -
DR EMBL: AC006922; AAM15324.1; -
DR EMBL: X62747; CAA44609.1; -
DR PIR: S52990; S52990.
DR HSSP: P21310; 1B8F.
DR InterPro: IPR008996; L-Aspartase-like.
DR InterPro: IPR001106; Phe/His_NH3-lyase.
DR InterPro: IPR005922; Phe_am_lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRPFAM: TIGR01226; phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
KM lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLNK 211 213 5-imidazolone (Ala-Gly)
FT (By similarity).
FT MOD_RES 212 212 DHA (2,3-DIHYDROALANINE)
FT (BY SIMILARITY).
FT CONFLICT 329 329 V -> I (IN REF. 2).
FT CONFLICT 426 426 R -> A (IN REF. 2).
FT CONFLICT 612 612 V -> I (IN REF. 2).
SQ SEQUENCE 725 AA; 78782 MW; 5D65F102D50A2946 CRC64;

Query Match 28.7%; Score 1035.5; DB 1; Length 725;
Best Local Similarity 36.4%; Pred. No. 1.3e-61;
Matches 271; Conservative 138; Mismatches 259; Indels 77; Gaps 21;

QY 18 NGSHAPTSAGPTBSAL-----RRPPGIDGHAARH--SQEIVQELISD 60
DB 4 NGAH-----KSNQGVADAMCGGDITKKNVITNAEDPLMWGAAEPMKSHLDEVKRVAAE 59
QY 61 PPTDVEELKGYSTIVDVVGAAKGRVYVQNDDEIRARVDSVDFLKAOLQNSV--YGV 118
DB 60 FRKPVNNGIGETITTCQVAAISTIGNSVKVELSEFARAGVANSWWSKXKTDSTYGV 119
QY 119 TTGFSGSADTRTDAVSLQKALIEHQLCGVPTXSXSSVSGRGENTLPLEVAGAVIR 178
DB 120 TTGFGATSRRTNGVALQKELIRFLNAGI-----FGSTKETSHTLPHSATRAAMIVR 172
QY 179 VNSLTGSHAVLVVLEALTNFLNRIPIYVPLRSGISASGSLSTSYIAAIIQHGDVX 238
DB 172

DB 173 INTLQPSGIRFELLBAITFELNNNTPSLPLRGTTSAGDVLPSYIAGLTGRNSK 232
QY 239 VVHLHGTE-KIMFARBAISLFGLEA--VLGPKRGGLVNGTAVASAMATLSHDSMTL 295
DB 233 A-----TGNGEALTBEEAFKLAGISSGFELQPKREGALVNGTAVSGKMSMVLFTNTVL 288
QY 296 SLISQALTLALVEANVGQGSFAPFIDVCRPHGQVEVARNITLISGSSFAVHEEEV 355
DB 289 SVLAETLAVFAVFNVSGRP-EFTDHLTRLKHHPQGLEAAVMEHILDDSSY-NKLQKL 346
QY 356 KYKDEGLIRQDRYLPRTSPQFLGFLVEDMMAYSTLSLENNTTDNPDLDEVKQTAHG 415
DB 347 HEMDELQPKQDRVALKRTSPQWLGQIEIVITVATKSRLEINSVNDNPLIDVSNKALIHG 406
QY 416 GNPQASAVSISMEKTRIALILGKINFTQCTELLNAAVRGLPSC-LAEDPSLNYHKGK 474
DB 407 GNPQCTPQGVSDNTRILAIRIGKLMFAQFSELVNDPNNLPSNLTAASNPSLDYFQK 466
QY 475 LDHIAAVASELGHANVTTFVPAEWMGNQAVNSLALISRRFAEANDVLSLASHLY 534
DB 467 AITANASTGSELQYLANVYTHVQSAEQHODVNSLGLISRKISEAVDILKMTSTFLV 526
QY 535 CTLQAVDLRAMELDFKQFDELPTLLQQLGTSLDVNAALAEVKALNKL-----EQ 588
DB 527 ALCQAVDLRHLLENLRQVTKVTSQVAKVLTG--VNG-ELHFSRRCEDKLLKVDREQ 583
QY 589 TTYDLEPRKMDAFYATGYVE-----LSSPSANVTLTAVNMKVASAKAI 638
DB 584 VYTVADDP-----CSATYPLIQIKRQYIVDHALVNGESEXNAVTSLFH-KIGAFEEEL 635
QY 639 S--LTREVR--NRFWQTSSQAPAAVYLSPTRTVYLSFVREELGVQARRGDVFGVQOET 694
DB 636 KAVLKEVEAARAAADNQTSAIP-NRIKEGRSYPLRYVRRELOT-----ELLTGEVYTS 689
QY 695 IGSNVSRIYEAIKGRINHVLVKKL 719
DB 690 PGEEDKVFYTAICEGKITDPMWECI 714

RESULT 6
PALL_PRUV
ID PALL_PRUV STANDARD; PRT; 717 AA.
AC 064963;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
GN PALL.
OS Prunus avium (Cherry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxID=42229;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Summit;
RA Wierema P.A., Wu Z.;
RT "A full-length cDNA for phenylalanine ammonia-lyase cloned from ripe
RT Sweet Cherry fruit (*Prunus avium*).";
RL (in) Plant Gene Register PCR98-184.
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF036948; AAC78457.1; -

DR HSSP: P21310; 1B8F.

DR InterPro: IPR008948; L-Asparagine-like.

DR InterPro: IPR001106; Phe/His NH3lyase.

DR InterPro: IPR005922; Phe_am_lyase.

DR Pfam: PF00221; PAL; 1.

DR TIGRfams: TIGR01226; phe_am_lyase; 1.

DR PROSITE: PS00488; PAL_HISTIDASE; 1.

KW Lyase; Phenylpropanoid metabolism; Multigene family.

FT CROSSLINK 203 205 5-imidazolone (Ala-Gly)

FT (By similarity)

FT MOD_RS 204 204 DHA (2,3-DIDEHYDROALANINE)

FT (By similarity)

FT NCBI: 717 AA; 77999 MW; B84DF90A0BF60B3 CRC64;

SEQUENCE

Query Match 28.4%; Score 1024; DB 1; Length 717;

Best Local Similarity 36.2%; Score No. 7.4e-61;

Matches 262; Conservative 133; Mismatches 269; Indels 60; Gaps 17;

14 NGFTNSHAAPTYSAGPTSLRRTP--GLDGHAAHSGLEIVGELLSDPTDVELISG 70

9 NGHNG-----SVLEPELCIKKDPPLNGVAETLKSHLDEVKMAEVRKPKVYKLG 61

71 YSLTVRVVGAARKGRVRYQNDDEIRAVDSVDFLKAQQLNSV--VGVTGGSGSDT 128

62 ESLTISQVAIAITHDSGVKVESASRAQVKSADVDWSKGTDSGVTTGFATSHR 121

129 RTEDAVSLQKALIEHOLCVPTXSXSFVGRGLENTLPLEVVGAMVIRVNSLTRGSA 188

122 RTQGAALQKELIRPLNAGVFGSTKES-----GHLPHQATRAMLVRINTLLQVSG 174

189 VRLVTEALTNFLNHTITPVPFRGISASGDLSPSTYAGKITGHDPVKVHLHEGTEK 248

175 IREILLEVITKPLNNVTPCLPRGTITASCGLVPLSYAGLWLTDRPNSKA-VGPDG--Q 231

249 IMPAREALISFGLA--VVIQPEKGLVNGTAVASMTSLSHSHSLSLSQLTALT 306

232 TISAAEAFEEVINGSGFELQPEKGLALNGTAVSGSLASTVLEPTNIALISLSXIF 291

307 VEVANVGQGSFAFPIHDVCRPHPGQVEVARNITLLSGSSFAVE---HEEEVVKDEG 362

292 AEVVGQKP-EFTDLTKLKHHPGQIEAAIMEHILDGSSYKAKKHEODPLQKP-- 347

363 ILMDORPLRTSPQIPLVEDMMHAYSTLSLENTTNDPLLDVENKQTHAGNFOASA 422

348 --KODRIALTSPQWGPQIETVIRISTKIEREIDSVNPNLIDSRKKAHGGNFQGT 405

423 VSI-SMEXTRIALALIKLNFTQCTELNANNGPSCIA-AEDPSLYNHGKGDINHIA 481

406 IGVSMDVTRLAISIGKLMFAPGSFELVDFYNNGLPSNLGGRNPSLDYGFKAIEAAS 465

482 YASELGLANPTTTFQPAEMNGQVANSIALISARTAEANDVSLLSLASHLXCLQAVD 541

466 YSELQFLANPVNHNQSAHONQDVNSIGLISSKRTAEAVDILMLMSTFLVALCOAID 525

542 LRAMELDFFKKQFDPLPTLLQOHLGTGLDVNALAE-VKALNKLLEO-----T 589

526 LRHLBENTRTVKNVTSQVAKRLLTGVVNGELHPSRFCEKDLKLVYDEEYFAYIDPCS 585

590 TTYDLEPRMHDAISYATGYVELLSSPSANTLTAVANAKYASAK-AIQLTEVARNRF 648

586 ATPLWQKLRQV-----VEHALTWGENEKNASTISFQKIYAFEEELKVLPLPEVDSAR 639

649 WQPPSGAPAHAYLSP-RTRVLYSPVREELGVQARGDPVGVQOETIGSVVSYYEAIK 707

640 AALDSSAGVPRNITECRSPLYKFRLELG-----AEVLTQEKVRSFGEECDKVFALIC 694

QY 708 DGR1 711

DB 695 ECK1 698

RESULT 7

PAL4 ARATH

ID PAL4 ARATH STANDARD; PRT; 707 AA.

AC G9545;

DT 16-OCT-2001 (Rel. 40. Created)

DT 16-OCT-2001 (Rel. 40. Last sequence update)

DT 10-OCT-2003 (Rel. 42. Last annotation update)

DE Probable phenylalanine ammonia-lyase (EC 4.3.1.5).

GN A73G10340 OR F14P13.6.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016720; PubMed=1110713;

RA Salacoubat M., Lemcke K., Rieger M., Ansoorge W., Unselid M.,

RA Fartman B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,

RA Deiseny M., Boutry M., Grivell L.A., Macho R., Pulgomech P.,

RA De Simone V., Choisme N., Artiguenave F., Robert C., Broillet P.,

RA Wucher P., Catolico L., Weisenbach J., Saurin W., Quetier F.,

RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Bernes V.,

RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,

RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,

RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,

RA Conrad A., Hornischer K., Kauer G., Loenert T.-H., Nordstiek G.,

RA Reichelt U., Scharte M., Schoen O., Barques M., Terol J., Climent J.,

RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,

RA Cooke R., Laude M., Berger-Liauro C., Purrelle B., Masny D.,

RA de Haan M., Maarse A.C., Alcaraz J.-P., Cortez A., Casacuberta E.,

RA Montori A., Argitrou A., Flores M., Liguori R., Vitale D.,

RA Mannheim G., Haase D., Schott H., Rudd S., Zaccaria J., Mewes H.-W.,

RA Mayer K.F.X., Kahl S., Town C.D., Koo H.L., Talion L.J., Jenkins J.,

RA Rooney T., Rizzo M., Walts A., Utechtack T., Fujii C.Y., Shea T.P.,

RA Creasy T.H., Haas B., Walts R., Wu D., Peterson J., Van Aken S.,

RA Pai G., Miltcher J., Sellers P., Gill J.E., Feldlyum T.V.,

RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,

RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,

RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,

RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,

RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,

RA Nakabe A., Yamada M., Yasuda M., Tabata S.;

RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis

thaliana".

RL Nature 408:820-822 (2000).

CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the

first reaction in the biosynthesis from L-phenylalanine of a wide

variety of natural products based on the phenylpropane skeleton.

CC - CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).

CC - PATHWAY: Phenylpropanoid biosynthesis; first step.

CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),

which is formed autocatalytically by cyclization and dehydration

of residues Ala-Ser-Gly (by similarity).

CC - SIMILARITY: Belongs to the PAL / histidase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

DR EMBL: AC009400; AAF02809.1; -

DR HSSP: P21310; 1B8F.

```

DR InterPro: IPR008948; L-Aspartase-like.
DR InterPro: IPR001106; Phe/His NH3lyase.
DR InterPro: IPR005922; Phe_am_lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRfam: TIGR01226; Phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
DR Kyr: Phenylpropanoid metabolism; Multigene family.
FT CROSSLNK 194 196 5-imidazolinone (Ala-Gly)
FT MOD_RES 195 195 (By similarity).
FT MOD_RES 195 195 DH4 (2,3-DIHYDROALANINE)
SQ SEQUENCE 707 AA; 76919 MW; B677430941B50B9B CRC64;

Query Match 28.3%; Score 1020; DB 1; Length 707;
Best Local Similarity 36.6%; Pred. No. 1,4e-60;
Matches 261; Conservative 116; Mismatches 248; Indels 88; Gaps 16;

QY 49 SCLEIYQELLSPDDELVEISGLTVDVGAAKGRVYQNDDEIRAVDKSVDFLK 108
DB 30 SHLDEYKRWYKRAVAVLQGETLITGVAAVARGGSGTVELAEARAVKSSWVM 89
QY 109 AQLQNSV--YGVTTGGGSAADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRG-LENT 165
DB 90 ESMNRGTDSYVTTGGATSHRRKGGALQNELRFLNAGI-----FGPGAGDTSHT 142
QY 166 LPLEVVRGAMVIRVNSLTREGSAVRLVLEALTNFLNRITPIYPLRGSSISASGDLSPLS 225
DB 143 LRPETRAMLVKVNLTQYSGIRFEIIEATIKLNEITPCLEPRGRTIASGDLVELS 202
QY 226 YIAGATGGHPDVKNVHLHSGTEKIMFAPRAISLFGLEAV--LGPREGIGLVNGTAVASM 284
DB 203 YIAGLTGSPNSA--VGPSG--ETTLASAPFLAVSSFFELQPEEGALVNGTAVASGL 259
QY 285 ATSLSDHMLSLSCALTALTVEAMVGGQGSFAFHIDVCRPHPGQVEVARNRTLSG 344
DB 260 ASVTLFDANILAVLSVMSAMFAEWQGR--EFTDHLTHKLNKHHGQLEAAIMEHILD 318
QY 345 SGPAAV---HEBEVKVYKDDDEGLRQDRYPLKTSQFGLPLVEDMMAVSTLSLENTTT 400
DB 319 SSYVKEAQLLHEVDPLQKP-----KODRVALTSPQWGLGPOLEVRATKMIIEEINSVN 373
QY 401 DNPDLVENKQTAHGGNFPASAVSISMEKTRALLIGKLNFTOCTELNANMRG;PSC 460
DB 374 DNPDLVSNRKALHGGNFGGTPIGVAMDSKRLAISIGKLNFAQSESLVNDPYNNGLPNS 433
QY 461 LA-AEDPSLNTGKGLDIIHAAVASSELGHLANPVTTFVQPAEMGNOAVNSDALISARRTA 519
DB 434 LSGRRPSLDYGRKGAEMIASGSELOFLANPVTNHQSAEQHNQDVNSLGLISSRRTA 493
QY 520 EANDVSLILASHLYCTQAVDLRAMELDFKQFPLPLTLLQCHLGTDVNAALAEVK 579
DB 494 EAVDILIKMTSTYLVALCOAVDLRHLLENKK-----AVK 528
QY 580 KALKKLEQTTTYDLEPRWHA-----FSYATG-----TV 609
DB 529 SANSQVAKVLTGANGELHPSRFTERDVLQVDEVEYFSTADDCSLTYPLMKLKHIL 588
QY 610 VELISSP--SANTVTLTAVNAKVASAERKALSTREV--RNRFQWTPSSQAPAHVLSFRT 666
DB 589 VDHALLADPEREANSATSVPHKICAFEAELKILLPREVERVREYEETGSAIANNIKCERS 648
QY 667 RVLYSVREELGQARGDPVFGVQGETIGSVSVSYEAIQGRNHLVWML 719
DB 649 YPLRYRVDLENT-----ELTGENVRSFGEEFPDKFLAISDGKIDPLELCL 696

RESULT 8
PALT TRISU STANDARD; PRT: 725 AA.
AC P45734;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```

```

DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PALL1.
OS Trifolium subterraneum (Subterranean clover).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Trifolioideae; Trifolium.
ON NCBI_Taxid:3900;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. Karridale; TISSUE-Leaf;
RX MEDLINE=94171061; PubMed=8125321;
RA Howles P.A., Artoli T., Weisman J.J.;
RT "Characterization of a phenylalanine ammonia-lyase multigene family
in Trifolium subterraneum.";
RL Gene 138:87-92(1994).
CC -!- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -!- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -!- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -!- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
DR EMBL: M91192; AAA1993.1; -.
DR HSSP: P21310; 1B8F.
DR InterPro: IPR008948; L-Aspartase-like.
DR InterPro: IPR001106; Phe/His NH3lyase.
DR InterPro: IPR005922; Phe_am_lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRfam: TIGR01226; Phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
DR Kyr: Phenylpropanoid metabolism; Multigene family.
FT CROSSLNK 211 213 5-imidazolinone (Ala-Gly)
FT MOD_RES 212 212 DH4 (2,3-DIHYDROALANINE)
FT MOD_RES 212 212 (By similarity).
SQ SEQUENCE 725 AA; 78998 MW; 70F9925C8622240 CRC64;

Query Match 28.2%; Score 1019; DB 1; Length 725;
Best Local Similarity 36.0%; Pred. No. 1.6e-60;
Matches 273; Conservative 127; Mismatches 254; Indels 104; Gaps 22;

QY 6 DSLATTIANGFTNGSHAPTSAGPTALRRTPGLDGHAAHQQLIEVQLSDPTDDV 65
DB 17 DDFCLTHAN--ANNKKV-----GVAAEAKKSGHLDVKNMVEYKRPV 65
QY 66 VELSGVSLTRVDVVG--AARKGRVRYQNDDEIRAVDKSVDFLQAQIQNSV--YGVTTGTF 122
DB 66 VDLGGFTLLISQVAAIAAHGATVEL--SESARAGVKAASSPWVESNMKKGTSVGVTTF 123
QY 123 GGSADRTEDAVSLQKALIEHQLCGVTPTSXSSFSVGRGLENLTPLEVRGAMVIRVNSL 182
DB 124 GATSHRRITQGGALQKELIRFLNAGI-----FGNGESNHTLPHRATRAAMVIRINTL 176
QY 183 TRGSAVRLVLEALTNFLNRITPIYPLRGSSISASGDLSPSYIAGATGHPDVKAVHL 242
DB 177 LQYSGIRFEIIEATIKLNNNITPCLEPRGRTIASGDLVELSYIAGLITGRSNSKA--- 233
QY 243 HGTETKIMAPRAISLFGLEA--VILGPKELGIVNGTAVASMAWTLSDHSWMLTSLSQ 300
DB 234 HPGSGMLNAKFAQLAGINABFFELQPKEGALVNGTAVASGLASTVLEPANTLAVASE 293

```

```

QY 301 ALATLVEANVGGQGSFAPIHDVCRPHQGVFAVNRITLLSGSPFAVE---HEEYVK 356
D 294 VLSAIFAEVWQGRP-EFTDHLTRKHLKHHGPOIEAAMEHIIHGSAVYADAKKHEMDPL 352
QY 357 VKODEGILRODERPLRTPSPQFLGPIVEDMMHAYSTLSLENNNTTNPPLDVENKOTAHG 416
D 353 QKP-----KODRYALRTPSPQFLGPIVEDMMHAYSTLSLENNNTTNPPLDVENKOTAHG 407
QY 417 NFOASAVSISMEXTRIALALIGKLNFTQCTELNANRGDSCA-AEDPSLNYHGKGL 475
D 408 NFOGTPIGSMNDTRIALALIGKLNFTQCTELNANRGDSCA-AEDPSLNYHGKGL 467
QY 476 DIHAAVASELGHANVTTFVQPAEMGNQAVNSLARSRTAANDVLSLASHYXC 535
D 468 EIMAAVSCSELQYLANVTTHVQSAEONHODVNSGLSSRTKKAIEILQMSSTFLIA 527
QY 536 TLQAVDIRAMELDFKQDFLPTLLQHLGTGLDVNALALEVKKALNKRLEQTTVDLE 595
D 528 LCGAIDLRHLEENLK-----NSVKNTVSQVAKKTLITIGVSGELH 566
QY 596 P-----RHMDAFSY-----ATGVVLEL-----SSPSANVTLT 624
D 567 PSFCECKDLKVVDRH-VFSYIDPCSATYPLAQKRLQVLDHALVNGESEKNSNTSIF 625
QY 625 AVNAMKVAASAEXAIS--LTREVEN-RFMQTPSSQAPAHAYLSFRTRVLYSPFVRELGVA 681
D 626 Q-----KATFEELKLLPKEVSAATAYENGNSITANKINCRCSPYLYKVBELGSL 681
QY 682 RRGDVEVVGQETGNSVSRIFYALIKGRINHLVLM 719
D 682 LTGRVY-----SPGECDKLLFTAMCGKIIDLPLKCL 714

```

RESULT: PETCR
ID PAL3 PETCR STANDARD; PRT; 718 AA.
AC P45729;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 3 (EC 4.3.1.5).
GN PAL3.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; asterids;
OC Campanulids; Apiales; Apiaceae; Apioideae; apioid superclade;
OC Apium clade; Petroselinum.
CX NCBI_TaxID=4043;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95010141; PubMed=7925471;
RA Appert C., Logemann E., Hahlbrock K., Schmid J., Amrhein N.;
RT "Structural and catalytic properties of the four phenylalanine
ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.)";
RL Eur. J. Biochem. 225:491-499(1994).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.

```

CC -----
DR EMBL; X81159; CAAB5057.1; -.
DR PIR; S48726; S48726.
DR HSSP; P21310; 1B8F.
DR InterPro; IPR008948; L-Asparagine-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am-lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfam; TIGR01226; phe_am-lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
FT Lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLINK 204 206 5-imidazolethione (Ala-Gly)
FT MOD_RES 205 205 (By similarity).
FT MOD_RES 205 205 DHA (2,3-DIDERIDROLANTINE)
FT SEQUENCE 718 AA; 78165 MW; 17786451FEDBC35 CRC64;
SQ
Query Match 28.2%; Score 1017.5; DB 1; Length 718;
Best Local Similarity 36.5%; Pred. No. 2e-60;
Matches 259; Conservative 128; Mismatches 275; Indels 65; Gaps 19;
QY 14 NGFTNGSAAPTASAGPTSLRTP---GLDGHAAHQSOLIVQELSDPTDDVYELS 69
D 5 NGFTNG-HA---NNGNGLDCKKEDPLNMGVAEALTGSHDEVKMAVEYKRPVYKLE 59
QY 70 GYSLIVRDVYG-AARKGRVRVONDDEIRAVDKSVDFPKAOLQNSV--VGVTTFGGSA 126
D 60 GETLITGVYAAISARDDGCVKVELSEKRAQVYKASSWVWDNMKKGDSIGVTTGGATS 119
QY 127 DRTEDAVSLQKALLIHQLCGVPTXSXSFVSGGLE---NTLPLEVVGAMVIRVNSLT 153
D 120 HRTKQAGALQKELIRFNAGI-----FSGAGAGANNLTLPASRTAAMLVRIYITLL 170
QY 184 RGHSAVLVYV.EALNPFNHRITPVLPGSISAGDLSPLSTAGATIGHPVKVHVL 243
D 171 QGYSGRFELIEAIKPFKNHNITPCLPRGTTASGDLVPLSTAGLTLRPSKSA--V 227
QY 244 EGTEKIMFAREAISLFGLEA--VVLGPNELGLVNGTAVASAMATLSLDSHMLSLISQA 301
D 228 GRTGVTLSPBEAFKLAGVGGFFELQPKREGALVNGTAVASGMAVYLPFANILAVLEV 287
QY 302 LIALIVEANVGGQGSFAPIHDVCRPHQGVFAVNRITLLSGSPFAVEHEEYVKADDE 361
D 288 MGAIFAEVWQGRP-EFTDHLTRKHLKHHGPOIEAAMEHIIHGSAVYADAKKHEMDPL 345
QY 362 GILRODERPLRTPSPQFLGPIVEDMMHAYSTLSLENNNTTNPPLDVENKOTAHGQFQAS 421
D 346 QKPQODRYALRTPSPQFLGPIVEDMMHAYSTLSLENNNTTNPPLDVENKOTAHGQFQAS 405
QY 481 AASLGHANVTTFVQPAEMGNQAVNSLARSRTAANDVLSLASHYCYTLQAV 540
D 466 STCSLQFLANVTTHVQSAEONHODVNSGLSSRTKKAIEILQMSSTFLVGLQAI 525
QY 541 DIRAMELDFKQDFLPTLLQHLGTGLDVNALALEVKKALNKRLEQ----- 568
D 526 DIRHLEENLKSVKNTVSQVAKKTLITIGVSGELH-VNG-ELHPSFCECKDLRIVDREIFAYID 582
QY 589 ---TTTYDLERPHMDAFSAYATGVVELSSPSANVTLTAVNAMKVAASAEXAIS--LTRE 643
D 583 DPCSATYPLMQKLR-----LTVEHALNNGDKERNLSTIRQKLAFAFDEIKALLPK 635
QY 644 VNN-RFMQTPSSQAPAHAYLSFRTRVLYSPFVRELGVAQARGVGVQOETGINSVSR 702
D 636 VETAAALAESGNPAIPNRIKRCRCSPYLYKVBELG-----EVLSEKTRSPGEEPEKY 690
QY 703 YEALIKGRINHLVLM 719
D 691 PTAMSKGEIIDLPLECL 707

```

```

RESULT 10
PALI TOBAC STANDARD; PRT; 715 AA.
ID PALI TOBAC
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
NCBI OR PALB.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Solanaceae; Nicotiana.
NCBI_TaxID=4097;
RX
RN SEQUENCE FROM N.A.
RA Fukasawa-Akeda T.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
TISUE=Callus;
TAGRAIN=cv. Bright yellow; TISUE=Callus;
RA Taguchi G., Sharan M., Gonda K., Yanagisawa K., Shimosaka M.,
RA Hayashida N., Okazaki M.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3) .
CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC - SIMILARITY: Belongs to the PAL / histidase family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for, commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch) .
CC
DR EMBL; M84466; AAA4122.1; -
DR EMBL; AB008200; BAA22948.1; -
DR PIR; S66343; S66343.
DR HSSP; P21310; 1B8F.
DR InterPro; IPR008948; L-Asparatase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfam; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00468; PAL_HISTIDASE; 1.
DR Lyase; Phenylpropanoid metabolism; Multigene family.
FT FT MOD_RES 201 203 5-imidazolone (Ala-Gly)
FT CROSSINX 201 203 (by similarity).
FT MOD_RES 202 202 DHA (2,3-DIHYDROALANINE)
FT FT (by similarity).
FT FT MOD_RES 202 202 DHA (2,3-DIHYDROALANINE)
FT FT (by similarity).
FT FT MOD_RES 202 202 DHA (2,3-DIHYDROALANINE)
FT FT (by similarity).
SQ SEQUENCE 715 AA; 77780 MW; 03CB4E8527394C62 CRC64;
Query Match 28.2%; Score 1016.5; DB 1; Length 715;
Best Local Similarity 36.0%; Pred. No. 2.4e-60;
Matches 270; Conservative 130; Mismatches 259; Indels 91; Gaps 19;
DB
QY 13 ANGFPTN-GSIAAATKSAAGPTSLARTPGLDGAHQSLGVQLSDPDDVLELSGY 71
DB 3 SNCHVNGEENFELCKRSADPLNWEKAESLRG----SHLDVKKQVSEFRKPKVKLQGE 57
QY 72 SLTVRVDVGA--RKGRRVYQNDDEIFARVDKSVDFIKQJQNSV--YGVTTFGGSA 126
DB 58 SLTVAGVVAALAVRDKRANGVKVKESEAAAGVAKASDWMGSMNKGDTGSYGVTTGFGATS 117

```

```

QY 127 DRTEDAVSLQKALIEHOLCVPTPSKSSSEVSGLENTLPLEVRRAMVIRNLSLRGH 186
DB 118 HRRITNGALOKELLRFPLNAGV-----FNGTETSHTEPLPSATRAMVIRINTLQGY 170
QY 187 SAVRLVLEALTNFNLHRIPTVPARGISASGDLSPSYAGAITPHDPVKVVLHEGT 246
DB 171 SGIRELEILEATKILNSWITPCLPRGITIASGDLVLSYAGILTRPNSKA-VGNG- 228
QY 247 EKIMPARPAISLPGHEA--VVLGPEGLGVNGVAVASNAATLSLHDSHMLSLSQALTA 304
DB 229 -ETLNAEEAFRAVGVNGGFPELOPREGLATVNGTAVSGMAVLPFSNILLAVSEVLTA 287
QY 305 LVEAMWGQGSFAFIHIDVGRPHGVAVARNRITLISGSSFAVEEERKVMDDGIL 364
DB 288 IFAEVMNGKP-EFTDLHYKIKHHPGQLEAAALVEHLDDSSY-VKAQQLHEMDPLQKP 345
QY 365 RODRYPLRTSPQFLGPELVEDMMAHYSTLSLENTTTPNPLDVENKQTAHGGNFQASVS 424
DB 346 KQDRVALRTSPQWGLQEVIRATKMTEREINSVNDNPILDVSRNKALRGNFQGPIG 405
QY 425 ISMEKTRIALALIGLNTQCTELLNAAMNGLPSC-LAAEDPSLNHGGLDHIHAYA 483
DB 406 VEMDARFALASIGLMPAQSELVNDYNNGLPSNLTASNPBLDYGFGAETAMASYC 465
QY 484 SELGLAMPVTFVQPAEMGQAVNSLALISARTAEANDVLSLLASHYCTLQAVDLR 543
DB 466 SELQPLAVPTNHVSABQHQDNVNSLGLISAKTAEVNDILKMSSTYVALQALDLR 525
QY 544 AMELDPKQFDPPLTLQOHGTGLDVNALALVKKALKLEQTTTYDLPEWMDA-- 601
DB 526 HLEENIK-----NA-----VKRTVSGVARTLTMGANGELHPARF 560
QY 602 -----PSYANG-----TVVELSSPSAVVTLTAVANAVAS 633
DB 561 CEKELLRIYDRYLPAYVDDPSCNYPIMQKROYLDVHAMNNESEKNNVSSIFQKIGA 620
QY 634 AE---KAISLREVEN-RFQWTPSQAPAHAYLSRRRVLYSFVREELGVARGDVAVG 689
DB 621 FDEELKAV-LPEVSAARALESNPALPNRITCRSPYLRFXKEIGT-----ELLTG 674
QY 690 VQGETIGSVSRIRAIKDGRIINHVLVKKL 719
DB 675 EKVRSPEGECDKVFTAMCNGQITDPMTECL 704

```

```

RESULT 11
PALY POPTR STANDARD; PRT; 715 AA.
ID PALY POPTR
AC P45730;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL.
OS Populus trichocarpa (Western balsam poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucoside I; Malpighiales; Salicaceae; Saliceae; Populus.
NCBI_TaxID=3694;
RX
RN SEQUENCE FROM N.A.
RA STRAIN=P. trichocarpa x P. deltoides; TISUE=leaf;
RA MEDLINE=94151434; PubMed=8108506;
RA Subramaniam R., Reinold S., Molitor E.K., Douglas C.J.;
RA "structure, inheritance, and expression of hybrid poplar (Populus
RA trichocarpa x Populus deltoides) phenylalanine ammonia-lyase genes.";
RL Plant Physiol. 102:71-83(1993).
CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3) .
CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).

```


CC -1- PPM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L11747; AAA33805.1; -
 DR HSSP: P21310; 1B8F.
 DR InterPro: IPR008948; L-Aspartase-like.
 DR InterPro: IPR001106; Phe/His_NH1lyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRPFAMs: TIGR01226; phe am lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 201 203 5-Imidazolimine (Ala-Gly)
 FT (By similarity).
 FT MOD_RES 202 202 DHA (2,3-DIHYDROXALANINE)
 FT (By similarity).
 FT SEQUENCE 715 AA; 77918 MW; 5B9E837A6E8AF2B6 CRC64;
 Query March 28.1%; Score 1016; DB 1; Length 715;
 Batch Local Similarity 35.8%; Pred. No. 2.5e-60;
 Matches 266; Conservative 132; Mismatches 276; Indels 68; Gaps 15;
 QY 8 LATTANGFTNGSHAAPTKSAAGPTAARRTPGLDGHAAHQSLVEVLSDPTDVAE 67
 DB 1 MEVTYKGYQNGSLSEICVNQRDPLSM-----GVAABAMKSGHLEDEVKRVADYRKKPVK 55
 QY 68 LSGYSLTVRYNG-AARKGRVRYVQNDDEIRARVDSYDPLKAGQGNV-YQVTTGGG 124
 DB 56 LSGETTLTAQVASIAHDTGDVAVELSESARGVASSWWDSDKQDTSYGVATTGGA 115
 QY 125 SADRTTEDAVSLQKALIEHLCGVPTFSXSSFSVGRGLE--NTDLEVRGAMVIRVNSL 182
 DB 116 TSHRRTRQGAGALCKELIRFLNAGI-----FNGTETCTTLPHSATRAAMLVIRNTL 166
 QY 183 TRGHSARVLYVEALNFINLNRITTPVPLRGSISASGDISPLSYAGAITGHPDYKVT 242
 DB 167 LQYSGIRFEILRLNNTNNTTCLPLRGTITASGDIVPSTYAGLLTGRPNVSKA-- 223
 QY 243 HEGTEKIMFARBAISLFGLEA--VYVGPKEGLGVNGTAVASAKVATLSHDSHMLSLSQ 300
 DB 224 TGTGTGVLDAAEFKAAAGIESGFELQPEKGLALVNGTAVSGLSMWLFETNVLAVLSE 283
 QY 301 ALTALTEAVANGQSGFAFPIHDVCRPHPGVEVANRTILVSSGFAVE---HEEEVK 356
 DB 284 LLSAIFAEVANGKRP-EFTDHLTRKLGKHPGQLEAAAIKEHILDSGAYKAKKLEHEDPL 342
 QY 357 VKDDEGLRDRYPLTSPQFLGPIVEMDMHAYVLSLENTTNDP-LDVENQQTAGG 416
 DB 343 QKP-----KODRFAITSPQWLGPOLEVRFSFKIEERINSVNNPILIVSRKKAHGG 397
 QY 417 NFOASAVISMEKTRIALALIGLNTCTCELINAMNRGLPSC-LAAEDSLVYHGKGL 475
 DB 398 NFGGTPLGVSMNVRLAISIGKLLPAGSELVNFYNGGLPSNLTABRNNSLDYGRGA 457
 QY 476 DIIIAVASELGLANPVTTFQOPAEHQANVSLALISARTAEANDVLSLLASHLYC 535
 DB 458 EIMASVSELYLAPVTTHVQSAEONQDVNSLGLISSRTEAVDILKMTSTPLVA 517
 QY 536 TLQAVDRAAMELDFKQDPPLLPTLLQCHLTGTLVNLALAEVKKALKRLEQ----- 588
 DB 518 LCAALDRHLEENLKSAVKNVYQSKVLTITGANGF---LHPSFCKELKLVYDREYV 574
 QY 589 -----TTTYDLPRWHDAPSYATGTVVELLSSSPSANVTLTAVNAKVASAKAIS- 639

DB 575 FAYVDDPCGATPELMQKRGV-----VDHALNGENGEKMFSTV-FQITEAREELKA 627
 QY 640 -LTREYRN-REFQCTSSQAPAHAYISPTRTVLVSFVREELGVARRGDVFGVQDETGS 697
 DB 628 LIPKEVESAPAAVDSGNSAIDNKIKECRSYPLXYFVREELGTV-----LITGERVSPGE 682
 QY 698 NVSRIVEATKDRIRHIVLVKWL 719
 DB 683 EFDKVFAMCGKIDIPVLECL 704
 RESULT 12
 PAL2_ARATH STANDARD; PRT; 717 AA.
 AC PAL2_ARATH
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
 GN PAL2 OR AT3G53260 OR T4D2.190.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salambouat M., Lemcke K., Rieger M., Ansoerge W., Unsel'd M.,
 RA Faltmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delany M., Boutry M., Grivell L.A., Maché R., Puigdomenech P.,
 RA De Simone V., Choisy N., Artiguenave F., Robert C., Broctier P.,
 RA Wincker P., Catolico L., Weissenbach J., Sautin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmach E., Dizonet H., Eifle H., Jordan N., Banger S.,
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 RA Reichelt J., Schaefer M., Schoen O., Barques M., Terol J., Clément J.,
 RA Navarro P., Collado C., Perez-Perez A., Olteneader B., Duchemin D.,
 RA Cooke R., Laurie M., Berger-Liauro C., Purrelle B., Maszy D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Montfort A., Argiriou A., Flores M., Ligouri R., Vitale D.,
 RA Manhandt G., Haase D., Schoof H., Rüd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Xoo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts R., Utecherback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Walts R., Wu D., Peterson J., Van Aken S.,
 RA Pail G., Miltscher J., Sellers P., Gill J.E., Feldlyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asanizu E.,
 RA Sasano S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Metanab A., Yamada M., Yasuda M., Tabata S.,
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RT thaliana".
 RL Nature 408:820-822(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim U., Dale J.M., Chen H., Shinn P., Palm C.J.,
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,

Db	226	-TGDDGSLAKENFEKAGISTGFPLOQPREGLATNGTAVGSGMAWVLFEANQAVLA	284
Qy	300	QALPALTVEANVGQGGSFAPFIHDVCRPRPGQVEYVARNTLLSSGSPVAVHEEVKVKD	359
Db	285	EVLSAIFAEVWVSGPR-EFTDHLTRLRKHHGQLEAAAIWEHLIDGSSY-MKLAQRYHEND	342
Qy	360	DEGLRQRRVPLRTSPQFLPVEDMMHAASLTLENNTTNTPLDVENKOTAHQGNFO	419
Db	343	PLQKPKQDRYLRKTSRPMWLGFPQIEVRQKTKSLERINRVNPNPLIDVRNKAHQGNFO	402
Qy	420	ASAVSISMEKTRALLALIGKLNFTQCTELINAAANRGLPSCI-AAEDPSLTHGGLDIH	478
Db	403	GLPIGVASNDNRNLRLAIIAGLMEFAQSESELVNDPNGLPSNTLASSNPGLDGFPGAELA	462
Qy	479	IAAVASEKGLHNAVTTTFVQPAEMGNQAVNSLLISARRTAANDVLSTLLASHLYCTLQ	538
Db	463	MASYSSEIQYIANVYTHVQSAEQHDVNSLGLISKRTSBAVDILKLMSTFLVGLICQ	522
Qy	539	AVDRLAMELDPKKQFDLPLTLLOQHLGTGLDVNALALEVKALKNLRL-----EQTTY	592
Db	523	AVDRLHEENLRQTVKNTVGSQVAKVLTGTINGE--LHPSFCEKDLIKVDRREQPTV	579
Qy	593	DLER--RKHDAFSATGTGYE-LLSSPSANVTLTVANMKVYASAEKALS-LTSEVNR	647
Db	580	VDDPESATPYLMQRLQGVIVDHAISNEETKNAVTSI-FQKIGAFEEELKAVLPKEVEA	638
Qy	648	FWQGPSSQAPAHAYLSP-RTRVLYSPFREELGVQARCGDFVVGQOETIGSNVSRILEAI	706
Db	639	RAAYNGNAPLENNIKCRRSPYLRFREELGKLTGKRV-----SPGEHDKVFTAM	693
Qy	707	KDGRINHYLVKML	719
Db	694	CEGKLIDPLMDCL	706

RESULT 13

PALY_DIGLA

AC 023924; STANDARD; PRT; 713 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Phenylalanine ammonia-lyase (EC 4.3.1.5).

OS Digitalis lanata (Foxglove).

OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophytes, Spermatophyta, Magnoliophyta, eudicotyledons, core eudicots, asterids, lamiales, Lamiales, Antirrhinaceae, Digitalideae, Digitalis.

NCBI_Taxid=493450;

RN [1]

RP SEQUENCE FROM N.A.

RA Thoeninger C.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBDJB databases.

CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the first reaction in the biosynthesis from L-phenylalanine of a wide variety of natural products based on the phenylpropane skeleton.

CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).

CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO), which is formed autocatalytically by cyclization and dehydration of residues Ala-Ser-Gly (by similarity).

CC -1- SIMILARITY: Belongs to the PAL / histidase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce> or send an email to license@isb-sib.ch).

CC

OR EMBL; AUTO02221; CA05251.1; --

DR HSSP; P21310; 188F.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His_NH1lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRPFAMS; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism; Multigene family.
 KW CROSSLNX 198 200 5-Imidazolinone (Ala-Gly)
 FT MOD_RES 199 199 (By similarity).
 FT FT DHA (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 713 AA; 77732 MW; 747C64C65BDC6A CRC64;
 Query Match 28.4%; Score 1013.5; DB 1; Length 713;
 Best Local Similarity 36.1%; Pred. No. 3.7e-60;
 Matches 256; Conservative 122; Mismatches 251; Indels 81; Gaps 14;
 QY 49 GQETVQELLSDPTDVELSGYSTLRDVGAARKGRVYQNDDEIRARYDKSVPLK 108
 DB 35 SHLDEVARVBEPRKTVKLGCELTITSQVAALAAARDNEVAVQLAESSRAGYKASDWM 94
 QY 109 AOLQNSV--YGVTTGGSGADRTEDAVSLQKALIEHQLCGVTPTSXSFVSGLENTL 166
 DB 95 ESMKRGTDYGVTTGFGATSHRTKQKGLAQKELREFLNAGI-----FNGTGSTHTL 147
 QY 167 PLEVVGAMVIRVNSLTRGSAVRLVLEALNPLNHRITPTVPLRGISASGDLSPASY 226
 DB 148 PHSATRAMLVRLNITLQSGIRFEILEITIKFLNHNITPCLPRKGTITASGDLVPLSY 207
 QY 227 IAGAITGHPVAVVHLHEGTEKMFARBAISLFGLEAVV--LGRREGILVNGTAVSAS 283
 DB 208 IAGILTRPNSKA--VGPNG--ESLNAEQAFKLAGNSGLFELQKKEALAVNGTAVGSG 264
 QY 284 MATSLSDSHMLSLTSCALTLALTEAMVGQGSFAPLHDVCRPHRGVEVARNRITLS 343
 DB 265 LAGIALYEANILSLLEWMSAVFAEVMGKP--EFTDHLTHKAKHPGQIEAAIWEHTLD 323
 QY 344 GSSFAVEHEEVKVDDEGLRDROYPLRTSPQLPVEDMMAVSTLSLENTTTDNP 403
 DB 324 GSSY--VYAAQKMEMLPQPKKODRYALRTSPQMLGPQEVIRATKMIEREINSVNDP 382
 QY 404 LIDVENKQTAHGNRPQASAVSISMEKTRIALILKTNFTCTELNANMRGIPSCLA- 462
 DB 383 LIDVSRKALHGNFPGTPIGVSMDSRLAISIKLMPAQSELVNFPYNNGLPSNLISG 442
 QY 463 AEDPSLNYHGKGLDIIIAAYASELGHLPVTFVQAPAMGQAVNSIALISARTKAN 522
 DB 443 GRNPSLDYGKGEIEMASVCELOFLANPVNHNQSAEQHNQDNQNSGLISSKITBAL 502
 QY 523 DVLSLLASHLYTTLOAVDLRAMELDFKQDPPLLPTLLQOHLGLGVNALLAEVKKAL 582
 DB 503 DILKMSSTYVALCAQIDLRLHE-----ENILRLSVNNTI 537
 QY 583 NKRLQOTTYYDLEPRHND-----FSY-----ATGTVA----- 611
 DB 538 SQYAKKITLTGVNGELHPSRCELDLLRVVDREYFAVDPQCSATYPLAKLQVLEH 597
 QY 612 -LISSPSANVTLTAVANAKVASAKAISLTREVN--RFWQTFSSQAPAAHAYLSPTRVL 669
 DB 598 ALKNGENENKASTSITQKIEAFELKAVLPKEVASAHVALEDEKPAIANRITCEGAYPL 657
 QY 670 YSPVREELGVQARGDVPFYVGQETIGSVNSIYALIDGRINNVLYML 719
 DB 658 YKIFREELGT-----NFLTEKVMSPGEBODVFAASKGLIVDPLKLCL 702

DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PAL.
 OS Camellia sinensis (Tea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Ericales; Theaceae; Camellia.
 OX NCBI_TaxID=4442;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Yabukita; TISSUE=Leaf;
 RA Matsumoto S., Takeuchi A., Hayatsu M., Kondo S.;
 RT Molecular cloning of phenylalanine ammonia-lyase cDNA and
 RT classification of varieties and cultivars of tea plants (Camellia
 RT sinensis) using the tea PAL cDNA probe.";
 RL Theor. Appl. Genet. 89:671-675(1994).
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC - PATHWAY: phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC - SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D26556; BAA05643.1; -.
 DR HSSP; P21310; 188F.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His_NH1lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRPFAMS; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase; Phenylpropanoid metabolism.
 KM CROSSLNX 200 202 5-Imidazolinone (Ala-Gly)
 FT MOD_RES 201 201 (By similarity).
 FT FT DHA (2,3-DIDEHYDROALANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 714 AA; 77751 MW; 5F0E7EBD0A89071 CRC64;
 Query Match 28.0%; Score 1012; DB 1; Length 714;
 Best Local Similarity 36.2%; Pred. No. 4.7e-60;
 Matches 268; Conservative 130; Mismatches 280; Indels 62; Gaps 18;
 QY 5 LDSLATTANFTGSHAAPTSAAGPTSLARRPGLDGHAAGSQETVQELLSDPTD 64
 DB 1 MDS-TTIAIGNV--GSGGSGPGFCIKDPLN-----GAAAMKSGHLEEVKGVVEERKP 52
 QY 65 VVELSGYSTLRDVGAARKGRVYQNDDEIRARYDKSVPLKAOQNSV--YGVTTG 122
 DB 53 VVRLGGEFTLITSQVAALAVNGSEVAVELSSARBGVYASSDWMESMNKGTDSYGVTTG 112
 QY 123 GGSADRTEDAVSLQKALIEHQLCGVTPTSXSFVSGLE--NTPLEVVGAMVIRV 180
 DB 113 GATSHRTKQKGLAQKELREFLNAGI-----FNGTGSTHTPQSATRAMLVRLN 163
 QY 181 SLTRGSAVRLVLEALNPLNHRITPTVPLRGISASGDLSPASYIAGITGHPVAVV 240
 DB 164 TLQGYSGIRFEILEAISKLPLNHNITPCLPRKGTITASGDLVPLSLTLAGILTRHNSKA- 222
 QY 241 VHLHEGTEKMFARBAISLFGLEA--VVLGRREGILVNGTAVSASAVTTLSDSHMLSL 298
 DB 223 --VGPTEGILAPKFAFLAGVEGGEFFELQKKEALAVNGTAVGSGGLASVLFANILAVL 280

```

QY 299 SQALTAALTEAVANGQGGSFAPFHIDVCPHPQGVAVANRITLLSSGSPFAVEHEEVKVK 358
DB 281 SEVLALAIPEAVNGQGP-BFTDHLTHKLHHPQGLEAAIMHEHIDGSSY-VKAAQKHEM 338
QY- 359 DDEGILRODRYPLRTSPQFLGPEVDMMAVSTSLNNNTTNDPLLDVENKQTAHGNF 418
DB 339 DLQKRPQORVRLRTSPQWGLIIVISSRTKIEREINSVNDPPLNVSNNKALHGNF 396
QY 419 QASAVSISNEKTRALLALIGKINFTQCTELLNMAANRGLPSCLA-AEDPSLNHYGKGLD 477
DB 399 QCTPIGVSMNDNRLVAVSIGKMFAPQFSELVNDPFYNNGLPSNLGGRNPSLDYFKKAEI 458
QY 478 HIAAVASELGHANVTTFVQPAEMGNQAVNSLALISARPAENDVLSLLASHLYCTL 537
DB 459 AWAAYGSELQFLANVTNVQSAEQHNDVNSGLISSRKTAEDVLIKMSSTYVALC 518
QY 538 QAVDILAMELDEKQFDPLLPILLQOHLGTGLDVNALALEVKALNRLEO----- 588
DB 519 QAVDLRHPENLRNTVSTVSQVAKRVLTMG--VNG-ELHPSRFCEKDLRVDRERYFA 575
QY 589 -----TTTYDEPRKMDAFSATGTVEVLLSSPSANVTLLANVAMVSAEKAIS--L 640
DB 576 YIDDPGSAVTPMLQKLRQ-----VIVEHALKNESEKLNLSITFOKIRAFEEIKTLL 628
QY 641 TREVERN-RFWQTPSSQAPAHAYLSPTRVLYSFVREEELGVQARRGDYFVGVOQETIGSNV 699
DB 629 PREVESTRAALINGNSAIPNRKKEGRSYPLYKFVAREELGT-----ELLTGKVASPGEEF 683
QY 700 SRIVEAIKDGRIINHVLVKML 719
DB 684 DKVFETALCKGEMIDPLMDCL 703

RESULT 15
PALI_SOYBN STANDARD; PRT; 713 AA.
ID P27991;
AC 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
PT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
PALI:
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OC NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. T225; TISSUE=leaf;
RX MEDLINE=92190550; PubMed=1799682;
RA Frank R.L., Vocklin L.O.;
RT "Sequence and structure of a phenylalanine ammonia-lyase gene from
RT Glycine max.";
RL DNA Seq. 1:335-346(1991).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
CC first reaction in the biosynthesis from L-phenylalanine of a wide
CC variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52953; CAA37129.1; -.
DB FIR; S22991; 922991.
DR HSSP; F21310; 1B8F.
DR InterPro; IPR008948; L-Asparase-like.
DR InterPro; IPR001106; phe/His_NH3lyase.
DR InterPro; IPR005922; phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLNK 199 201
FT MOD_RES 200 200
FT FT DHA (2,3-DIDEHYDROALANINE)
SQ SEQUENCE 713 AA; 77744 MW; 9D71E1CC230216A CRC64;

Query Match 28.0%; Score 1011; DB 1; Length 713;
Best Local Similarity 36.9%; Pred. No. 5.5e-60;
Matches 270; Conservative 120; Mismatches 282; Indels 60; Gaps 19;

QY 14 NGFTGSHAAPTYSAGPTSLRRTPGLDGHAAHQSEIYQCELLSPDTPDVELSEYSL 73
DB 5 NGHQGSPCLST-AKGNNDPLNM-GAALAMGSHIDYKRVAVAEYKRVARLGGETL 60
QY 74 TVRDVGAARKGRVRVQNDDEIRARVDKSVDFLKAOLQNSV-YGVTTGFGSADTRTE 131
DB 61 TIAQVAAVAGHDGVAVELSESAREGVKASSEMVMNSMNGTDSYGVTGGCAISHRTK 120
QY 132 DAVSLQKALIEHQLCGYPTPTXSSFSVGRGLENTLPLEVNGAVNIVNSLTRGSAVRL 191
DB 121 QGGALEKELIRFLNAGI-----FGNGTSSHTLPHTATPAALVRLNTLLQGSIRF 173
QY 192 VVLEALTFNLRIRIPVIRLGRSISGSLSPSLYAGAITGHPDVKXVLAHEGTEKIMF 251
DB 174 ELLEALTKLNNNTVPCDLNGRTIASGDVPLSYIAGLLGRNSKA-VQPSG-EVLN 230
QY 252 ARBAISLFGEL-EAVVLGPKREGGLVNGTANVASANATLSLDSHMLSLISQALTAALTEA 309
DB 231 AKAEAFELASINSEFELQPKEGALVNGTAVGSLASVWLEAMITLAVLSVLAIPEAV 290
QY 310 MYGQGGSFAPFHIDVCPHPQGVAVANRITLLSSGSPFAVE-----HEBEVVKDDEGILR 365
DB 291 MQGKP-EFTDHLTHKLHHPQGLEAAIMHEHIDGSSYVKAAXKLHELDPLQKP-----K 344
QY 366 QDRYPLRTSPQFLGPEVDMMAVSTSLNNNTTNDPLLDVENKQTAHGNPQASAVSI 425
DB 345 QDRYALRTSPQWGLIIVISSRTKIEREINSVNDPPLNVSNNKALHGNFQCTPIGV 404
QY 426 SMEKTRLALALIGKINFTQCTELLNMAANRGLPSCLA-AEDPSLNHYGKGDHIAAYAS 484
DB 405 SMDNTRIALASIGKMFAPFSELVNDPFYNNGLPSNLTSRNPSSLDYGFKAELIAYASICS 464
QY 485 ELGHLANPYTTFVQPAEMGNQAVNSLALISARPAENDVLSLLASHLYCTLQAVDLRA 544
DB 465 ELQYLANPPTTHVQSAEQHNDVNSGLISSRKTEAELIKMSSTFLALCQALDLRH 524
QY 545 MEIDFKQGFDPPLLPILLQOHLGTGLDVNALALEVKALNKLEQTTTDLBRPHDASYS 604
DB 525 LEEHLKNSVKNVVSQVSKILLTG--VNG-ELHPSRFCEKDLRVDRERY-----ISY 575
QY 605 ---ATGTVVELLSSPSANVTLLAVNA-----MKVASAEKAIS--LTREVRNR 648
DB 576 IDPPGSAVTPMLQKLRQVLDALVNAECEKVNSSITOKLAIFEEELKNLLPKVEGAR 635
QY 649 WQTPSSQAPAHAYLSR-RTRVLYSFVREEELGVQARRGDYFVGVOQETIGSNVRIYEAIK 707
DB 636 AAYESKKAALPKKIQCRSYPLYKFVAREELGT-----GLLTGKVASPGEEFKDLFTAMC 690
QY 708 DGRINHVLVKML 719
DB 691 QGRKIDPLMECL 702

```

Mon Sep 13 10:31:12 2004

us-09-939-408a-13.rsp

Page 14

Search completed: September 9, 2004, 10:04:39
Job time : 19.4315 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:57:28 ; Search time 87.6349 Seconds
(without alignments)
2592.269 Million cell updates/sec

Title: US-09-939-408A-13
Perfect score: 3610
Sequence: 1 MAPSIDSIATITLNGFTNGS.....RIYEAIKGRINHVLVLMGLA 720

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_vxtn:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1041.5	28.9	725	10 Q94ANI	Q94ANI arabidopsis
2	1031.5	28.6	730	10 Q9W567	Q9W567 rubus idaeu
3	1025	28.4	711	10 Q8W2E4	Q8W2E4 lactuca sat
4	1018	28.2	711	10 Q9AX15	Q9AX15 pharbitis n
5	1016.5	28.2	707	10 Q7X7G3	Q7X7G3 oryza sativ
6	1016.5	28.2	717	10 Q8WMP4	Q8WMP4 arabidopsis
7	1014.5	28.1	708	10 Q84ENC	Q84ENC rehmannia g
8	1013	28.1	715	10 Q852S1	Q852S1 daucus caro
9	1010.5	28.0	714	10 Q8H6V6	Q8H6V6 populus tre
10	1007.5	27.9	714	10 Q7X6M8	Q7X6M8 populus tre
11	1003.5	27.8	671	10 Q84VE0	Q84VE0 oryza sativ
12	1002.5	27.8	721	10 Q9XFX5	Q9XFX5 citrus clem
13	998	27.6	713	10 Q8GZR8	Q8GZR8 lactuca sat
14	997.5	27.6	703	10 Q8YXG7	Q8YXG7 zea mays (m
15	997	27.6	716	10 Q9ATN7	Q9ATN7 asaccharie r
16	994	27.5	717	10 Q8H6W0	Q8H6W0 coffea cane

17	991.5	27.5	716	10 Q9MAX1	Q9MAX1 catharanthu
18	990.5	27.4	710	10 Q94C45	Q94C45 manihot esc
19	990	27.4	715	10 Q24266	Q24266 populus kit
20	984.5	27.3	687	10 Q94EP0	Q94EP0 manihot esc
21	982	27.2	711	10 Q8H6V5	Q8H6V5 populus tre
22	980.5	27.2	712	10 Q94F89	Q94F89 manihot esc
23	971.5	26.9	655	10 Q9XGR3	Q9XGR3 vigna ungui
24	967.5	26.8	718	10 Q9XFX6	Q9XFX6 citrus clem
25	957.5	26.5	703	10 Q84L12	Q84L12 phalaenopsi
26	957	26.5	681	10 Q8RUZ3	Q8RUZ3 pinus sylve
27	956	26.5	681	10 Q8SAS6	Q8SAS6 pinus sylve
28	956	26.5	681	10 Q8RY49	Q8RY49 pinus sylve
29	955	26.5	549	10 Q04876	Q04876 hordeum vul
30	950	26.3	681	10 Q8SAS7	Q8SAS7 pinus sylve
31	946	26.2	619	10 Q39545	Q39545 cucumis mel
32	936	25.9	619	10 Q8H6V9	Q8H6V9 coffea cane
33	934	25.9	618	10 Q9FS00	Q9FS00 dianthus ca
34	930.5	25.8	727	10 Q7XJC3	Q7XJC3 pinus pinsa
35	929.5	25.7	727	10 Q7XJC4	Q7XJC4 pinus pinsa
36	786	21.8	396	10 Q8WMP6	Q8WMP6 beta vulgar
37	780.5	21.6	532	10 Q81E47	Q81E47 capsicum ch
38	769.5	21.3	544	10 Q9FT38	Q9FT38 rubus idaeu
39	741.5	20.5	362	10 Q9FUI7	Q9FUI7 rubus idaeu
40	741.5	20.5	380	10 Q84U27	Q84U27 ginkgo bil
41	714.5	19.8	363	10 Q9FUI8	Q9FUI8 rubus idaeu
42	659.5	18.3	497	10 Q04869	Q04869 hordeum vul
43	652.5	18.2	475	10 Q04875	Q04875 hordeum vul
44	620	17.2	435	10 Q9M5U1	Q9M5U1 betula verr
45	619	17.1	436	10 Q9LEP1	Q9LEP1 betula verr

ALIGNMENTS

RESULT 1
ID Q94ANI PRELIMINARY; PRT; 725 AA.
AC Q94ANI;
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia lyase (Hypothetical protein) (BC 4.3.1.5).
GN A12G37040 OR PAL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroidae II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koeseima E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narasaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinzaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full length cDNA of gene T1J8-22/A12G37040 (GI:6598547).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,
RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamuya A.,
RA Kawai J., Kim C., Narasaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA Kato M., Seki M., Shinn P., Tang C.C., Toriumi M., Wong C., Wu H.C.,
RA Davis R.W.,
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Ergu P., Lee J.M.,

RA Tortum M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RY
 RP SEQUENCE FROM N.A.
 RA Cochran F.C., Kim M.R., Cardenas C.L., Davin L.B., Lewis N.G.,
 RT "Putative PAL multigene family in Arabidopsis thaliana: Kinetic
 RT Characterization of Recombinant Proteins and Knock-out Mutant
 RT Characterization.";
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY045919; AK76593.1; -
 DR EMBL; BT003350; AA029949.1; -
 DR EMBL; AY079363; AAL85094.1; -
 DR EMBL; AY303128; AAP59438.1; -
 DR PIR; G84787; G84787.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia ligase activity; IEA.
 DR GO; GO:0016841; F:ammonia-lyase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMS; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Hypothetical protein; Lyase.
 SQ SEQUENCE 725 AA; 78725 MW; 026263B2DDEFE9CE CRC64;
 Query Match 28.3%; Score 1041.5; DB 10; Length 725;
 Best Local Similarity 36.5%; Pred. No. 3.2e-61;
 Matches 272; Conservative 138; Mismatches 258; Indels 77; Gaps 21;
 QY 18 NSHAAPTKSAAGPTSL-----RRTFGLDGHAHQ---SOLEIYQELLSD 60
 DB 4 NGAH---KSNGGVAMLCGDIDIKKNVINAEDPLWGAALAEQMSKSHLDEKRWAE 59
 QY 61 PTDDVYELSGYSLTRVDVVGARKGRVRVQNDDEIRARVDKSVFLAQLQNSV--YGV 118
 DB 60 FRKPVNLGGETLTICQVAALSTIGNSVYVELSETPARAGVNASDPWVESNMKGTDSGV 119
 QY 119 TTGFGSADRTEDAVSLQKALIEHQLCVPTSSFSVRGLENLPLEVVGANVIR 178
 DB 120 TTGFGATSHRTKNGVALQKELIRPLNAGI-----RSTKETSHLPHSATPAALVR 172
 QY 179 VNSLTGHSAAVRLVLEALTNFLNRHTPIVRLGSGISASGDLPSYIAGATGHPDVX 238
 DB 173 INTLGGFSGIRFEIIEATISFLNNNITPSLPLRGTITASGDVPLSYIAGLIGRPNSK 232
 QY 239 VHLHNGTE-KIMFAEALSLFGLA--VILGPKGGLGVNCTRAVSAMATLSHDSML 295
 DB 233 A---TGNGEALTAFAEAKLAGISSGFEDLPQKGLLVNGTAGSGMSAMVLFETNVL 288
 QY 296 SLISQALTAALTEAMVGGQGSFAPFIHVCSPHPQOVAVANIRITLLSGSFAVEHEEV 355
 DB 289 SVLAETLSAVFAEVNSGKP-EFTDHLTRLKHHPQIEAALIMEHIDGSSY-WKLAQKL 346
 QY 356 KVVCDDEGILRODRYPIRTSPQFLGVLEDMMHAYSTLSLENNTTTNDPLDVENKQYAA 415
 DB 347 HEMDPLQKPKQRYALRTSPQWLGGQIEVIRATSTSIEREINSVNDPLDIVSRKAHG 406
 QY 416 GNFOAGAVSISMEKRLATLALIGKLNFTQCTELNAANRGJPSCL-AAEDPSLNYGKG 474
 DB 407 GNFQGTPIGVSMNDRLALAIAGKLMFQFSELVVDVFNNGLPSNLNRSRPSLDYFPGK 466
 QY 475 LDIHIAVASELGHLANPVTTVQPAENGQAVNSLALISARTPAEANDVLSLLASHLY 534
 DB 467 AELAMASYSELOYLANPVTHVQGAEOHNDVNSGLISSRKTSEAVDILKIMSTFVL 526
 QY 535 CTLQAVDLRAVELDFKQDPDLLPTLLQGHGTLGDVVALALEVKKALNKL-----EQ 588

DB 527 AICQAVDLHLHLENIRQTVKNTVSGVAKKVLTTG--VNG-ELHSPRFCEKOLLKYVDREQ 583
 QY 589 TTTYDLEPRMHAFASTACTIVE-----LSSSPSANVLTIVANAKVASAKAI 638
 DB 584 VYTVADP-----CSATYPLIQKLVQVVDHALNGSEKNAVTSIFH--KIGAFEEEL 635
 QY 639 S--LTREYA--NRFWQTSQAPAHAYLSPRTVLYSFVEBELGVQARCDVPFGVQOET 694
 DB 636 KAVLPREVGAARAAVNDNGTSALP-NRIKCGSYPLRYRREBELGT-----ELLTGEKVT 669
 QY 695 IGSNVSRIYEAIKDGRINHLVYKML 719
 DB 690 PGEPEFDKVTATCEGKIIDPMWECI 714
 RESULT 2
 ID Q9NS67 PRELIMINARY; PRT; 730 AA.
 AC Q9NS67;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
 GN PAL2.
 OS Rubus idaeus (Raspberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurossids I; Rosales; Rosaceae; Rosoideae; Rubus.
 OX NCBI_TaxID=32247;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21437948; PubMed=11553751;
 RA Kumar A., Ellis B.E.;
 RT "The phenylalanine ammonia-lyase gene family in raspberry, structure,
 RT expression, and evolution.";
 RL Plant Physiol. 127:230-239(2001).
 DR EMBL; AF237955; AAP40224.1; -
 DR HSBP; P21310; 188F.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia ligase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
 DR GO; GO:0009058; P:biosynthesis; IEA.
 DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMS; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 730 AA; 79422 MW; 0E942417E50CE113 CRC64;
 Query Match 28.6%; Score 1031.5; DB 10; Length 730;
 Best Local Similarity 37.1%; Pred. No. 1.5e-60;
 Matches 268; Conservative 130; Mismatches 276; Indels 49; Gaps 17;
 QY 14 NGFTNGS--HAAPKSAAGPTSLARTRFGLDGHAHQ---SOLEIYQELLSDPTDVVEL 68
 DB 13 NGIQGSLDDGLCTIKTSIKTGYSDPLNMWGAALAEBSMGSHLDEVRAVAEKRKPVNL 72
 QY 69 SGYSLTVDVVGAAARKRRVRVQNDDEIRARVDKSVFLAQLQNSV--YGVTTGGGSA 126
 DB 73 GGETITLISQVAALINNHSGVYVELAEASRRAGVASSDPWVDSNMKGTDSYGVTTGGCATS 132
 QY 127 DTRTEDAVSLQKALIEHQLCVPTSSFSVRGLENLPLEVVGANVIRVNSLTGCH 186
 DB 133 HRTKQGAALQKELIRPLNAGVLANGTES-----AHTLPHSATPAALVIRINTLQGY 185
 QY 187 SAVLVVLEALTNFLNRHTPIVRLGSGISASGDLPSYIAGATGHPDVGVVHVEGT 246
 DB 186 SGIRFEIIEALSKLNINNTPLCPKGTITASGDVPLSYIAGLIGRPNSKA-VGPKG- 243

Query Match	28.4%	Score 1025	DB 10	Length 711
Best Local Similarity	37.9%	Pred. No. 4e-60		
Matches	267	Conservative 127	Mismatches 259	Indels 52
Gaps	18			
QY	40	GLDGHAAHQSOLEIYQELLSPDIDDVYELSGSYLTRYADVVG--AAKGRRVQNDZIR	97	
DB	23	GAAAEHLTGSHLDEYKCVAAFRKRPVKLGGELLITYSQAAGLAANDSTVYELSEAR	82	
QY	98	ARYDKSVDFLKAQLONSV--YGVTTGGSGADTRTEDVASLQKALIEHQLCGVPTYSXS	155	
DB	83	AGVKAASDVMWESNMKKJDSYGVTTGGATSHRRTKGGALQKELIRFNMAGI-----	135	
QY	156	FVSGRLGENTLPLEYVRGAMVIRNLSLRGHSANVLVLEALNPNHRITLVLPGSI	215	
DB	136	FNGETSTHTPLSHSATRKAMTVRINTLLQSGISGRFEILATKFLNNVITCLPRLGTT	195	
QY	216	SASGDISPLSYAGAITGHPDVKHHVLEHGTEKIMFARBAISLFGLEA--VVLGPKEGIG	273	
DB	196	TASGDVLPLSTIAGLLTRPKSKA---VGPTEYVLAERKFAAAGVEGGFELQREGIA	252	
QY	274	LUNGTVASASMATLSLHDSHMLSLSQALLNLYEAMVQGGSPAFFIHDVCRPHGYVE	333	
DB	253	LVNGTVGSGMASMVLFPANVLTALLSEVLSAIFAEVWQGR-FETDHLTRKLLKHNGQLE	311	
QY	334	VAANIPTLLSGSSFAVEHEEYKVKYKDEGILRQDRYPILTSPOPLVIEDMMHAYSTLS	393	
DB	312	AAAIWYIILDSGY-VKAAQKVHEMDPLQPKQDRYALRTSPQMLQPIEVRISSTXME	370	
QY	394	LENNITTDNPLLDVENKQTAHGQNFQASAVISIMKTRTLALIGLNTFOCTELINAM	453	
DB	371	REINSVNDNPLLDVARNKALHGNFQGPICVSDNTRLLAIGKLMFAQPSSELVNDFY	430	
QY	454	NRLPSPCLA-AADPSLNVHKGGLDHLIAAYASELGHANVTTFVQPAEKGQNAVSLV	512	
DB	431	NNQLPBNLISGGRNPLDYGFKGGIYANASYSSELQPLANVYTNHVSALBNHDVSLGL	490	
QY	513	ISARRFAENADVLSLLASHLYCTLQAVDLFAMELDFKQFDPLLFTLLQOHLGTGLDVN	572	
DB	491	ISARKTAEFADITKLMSFTYVALQOSIDLHLEENKSTYKATVQSAKAVLTWG--VN	548	
QY	573	ALALEYKALKNLRLQTTTYDLERPMHAFSY---ATGR-----VVELLSSSP	617	
DB	549	G-ELHSPRCEKDXLLRYVDREY-----VFAYIDVCSGYPLMQIKROYLDHALNNEE	601	
QY	618	SANVTTLTANAMKVASAEKALS--LTREVRN-RFQTPSSQAPAHVYLSPTREVLVSFVR	674	
DB	602	TEKNTTST-FQKIATFEELKVLPLKVEGVRLAYENDLSTIPNHIKAKRSYPLVRFVR	660	
QY	675	EELGVARRGDVVFVGVOETTIGSNVSRIEAKIKDGRINHVLTQML	719	
DB	661	EELG----RG-FLTGKVTSPGEERDFVFTAMCKGOIIDBLECL	700	
RESULT 4				
Q9AXIS				
ID	Q9AXIS	PRELIMINARY:	PRT:	711 AA.
AC	Q9AXIS			
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DS	Phenylalanine ammonia-lyase (EC 4.3.1.5).			
OS	Pharbitis nil (Violet) (Japanese morning glory).			
OC	Eukaryota; Viridiplanetae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; asterids;			
OC	lamids; Solanales; Convolvulaceae; Ipomoea.			
OX	NCBI_TaxID=35883;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Violet;			
RA	Nakazawa A., Nozue M., Yasuda H., Takeba G., Kubo H.;			
RT	"Expression of phenylalanine ammonia-lyase in Pharbitis nil.";			
RL	Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.			


```

Db 576 ANPLMTKIRAVLVEHALANGPAEKDSSVHS-KITAFEEELREALPREMEARVAFT 634
Qy 654 SQAP-AHAYLSPTRYLVSFVEELGVQARRGDVGVQOETIGSNVRIEAIKDRIN 712
Db 635 GAAPIRNRKERSRFLVYFVAREELGCV-----YITGEKLSPGEECKVFLAISEKRLI 689
Qy 713 HVLVKKL 719
Db 690 DPMLECL 696

RESULT 6
Q8RMP4 PRELIMINARY; PRT; 717 AA.
ID Q8RMP4 AC
Q8RMP4 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase.
AT3G53260.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxId=3702;
OX
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shimozaki K.,
RA Ecker J., Theologis A., Davis R.W.,
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shimozaki K.,
RA Ecker J., Theologis A., Davis R.W.,
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY092957; AAM12956.1; -
DR EMBL: BT000035; AAM15354.1; -
DR GO: GO:0005737; C-cytoplasm; IEA.
DR GO: GO:0016211; F-ammonia ligase activity; IEA.
DR GO: GO:0016841; F-ammonia-lyase activity; IEA.
DR GO: GO:0016829; F-lyase activity; IEA.
DR GO: GO:0009058; P-biosynthesis; IEA.
DR GO: GO:0006559; P-phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Aspartase-like.
DR InterPro: IPR001106; Phe/His NH3-lyase.
DR InterPro: IPR005922; Phe_am-lyase.
DR Pfam: PF00221; PAL; 1.
DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
KM Lyase.
SQ SEQUENCE 717 AA; 77801 MW; C4C9CBDCB91A838 CRC64;

Query Match 28.2%; Score 1016.5; DB 10; Length 717;
Best Local Similarity 36.0%; Pred. No. 1.5e-59;
Matches 264; Conservative 141; Mismatches 283; Indels 45; Gaps 17;

Db 1 MDJEMLCGGGKRTVAVTTKLADPLNW-----GLADQKXSHDEVKMAAEIRP 55
Qy 65 VVELSGVSLVRDVGARKGRVRVQNDDEIRAVDKSVDFLKAQLONSV--YGVTTGF 122
Db 56 VVNLGSELTILIGVAAISTVGVKVELAETSRAGVYKASDPMVMSNMKGTDSYVTTGF 115
Qy 123 GGSADRTEDAVSLQALLIEHQLCGVTPISSXSFVSGLENTLPLEYVRGAMVIRNVL 182

```

```

Db 116 GATSHRRRTKNGTALOTELIRFLNAGI-----FGNTKETCHTLPOGATRAMLVNVL 168
Qy 183 TGHSAVRLVVEALTNPLNRTPIVPLRGSISASGDLPSLSTAGATGHPVVKVHL 242
Db 169 LOGYGINFEILAEATLSLNNNISPSLPLRGTITASGDLVPLSLIAGILICRPSKA--- 225
Qy 243 HEGTE-KIMFAEALSLFGLA--VLGPKGGLGVNGTAVASAMATLSLHDSMILLS 299
Db 226 -TGPDGESLTAKAEAEKAGISTGFEDLPQKEGLALVNGTAVAGSGMAVMVFEANQAVLA 284
Qy 306 QALTALTYEAMVQGGGFPAPFLHDVCRPHGVGVANRIRITLSSGSPFAVEHEEVKVD 359
Db 285 EYLSAIFAEVNSGR-EFTDHLTRLRKHHPQIAPALMEILLGSSY-MTLAQKHEMD 342
Qy 360 DGLIRQDRYPLRTSPQELGVLVEDMMAVSTLSLENTTNDPLDVENKQTAHGNFQ 419
Db 343 PLOKRODPRVALRTSPQELGVLVEDMMAVSTLSLENTTNDPLDVENKQTAHGNFQ 402
Qy 420 ASAVSISEKTRIALALGKLNFTQCTELLNAAANRGLPSCL-AAEPPSLNHGKGLDIH 478
Db 403 GPPIGSMNDRLALIAALGKMFQFSSLVNDFYNNGLPSVLTASNPISLDYGRKGAELA 462
Qy 479 IAAVSELGHLNPTTTPQAPAEKNGQAVNSLALISARTAEANDVLSLASHLYCTLQ 538
Db 463 MASYSSELQYLANPVTSHVQAGEHNDQVNSLGLISSKRTSEAVDILKMTSTFLVGICQ 522
Qy 539 AYDRLAMELDFKQDPLPTLQOHLGTGLDVNALAEVKALANKU-----EQTTY 592
Db 523 AYDLHLLENNLRQYKXNVSOVAKKLVTTGNGE---LHPSRFCEKLLKVDREOVFTY 579
Qy 593 DLEP--RWEADASVATGTVVE-LSSSPSANVTLTAVANMVASAEKALS-LTREVRNR 647
Db 580 VDDPSAITYPLMQRROYIVDHALSNGETEKNAVTSI-FQKIGAFEBELKAVLKEVEEA 638
Qy 648 FQWTPSSQAPAHAYLSP-RIRVLYSFVEELGVQARRGDVGVQOETIGSNVRIEAI 706
Db 639 RAAYNGTAPIPNRKEXRSPVLYFVAREELGKLTNGEXV-----SPGEFPDVFETAM 693
Qy 707 KQGRINHLVKKL 719
Db 694 CEKGLDPLMDCL 706

RESULT 7
Q94ENO PRELIMINARY; PRT; 708 AA.
ID Q94ENO AC
Q94ENO 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL1.
OS Rehmannia glutinosa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Lamiales incertae sedis; Rehmannia.
OX NCBI_TaxId=99300;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. Keumtan;
RA Yun S.J., Lee B.K., Park M.R., Srinivas B., Chun J.-C.;
RL "Molecular characterization of phenylalanine ammonia-lyase gene in
RL Rehmannia glutinosa."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF401636; AAK84225.1; -
DR GO: GO:0005737; C-cytoplasm; IEA.
DR GO: GO:0016211; F-ammonia ligase activity; IEA.
DR GO: GO:0016829; F-lyase activity; IEA.
DR GO: GO:0045548; F-phenylalanine ammonia-lyase activity; IEA.
DR GO: GO:0009058; P-biosynthesis; IEA.
DR GO: GO:0006559; P-phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Aspartase-like.

```

DR InterPro: IPR001106; Phe/His NHase.
DR InterPro: IPR005922; Phe am lyase.
DR InterPro: IPR006162; Prantase_S.
DR Pfam: PF00221; PAL_1.
DR TIGRfams: TIGR01226; phe.am.lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
KW Lyase.
SEQUENCE 708 AA; 76835 MW; 1B6AF08816F1B8FE CRC64;

Query Match 28.1%; Score 1014.5; DB 10; Length 708;
Best Local Similarity 35.6%; Pred. No. 2e-59;

Matches 264; Conservative 122; Mismatches 269; Indels 87; Gaps 17;

DR 18 NGSHAAP---TKSAGPTSLRRTPLDGHAAHQSLEIVQELLDPPDDVVELSGYSLT 74
DB 3 NGHHSNGLCVETTRPLMNVAAESLKG-----SHLDVKRMVEEFKRPVKLGESLT 57
QY 75 VRDVGAARKGRVRYONDEIRARYDKSVDFLKQLONSV--YGVTTGGGSADRTED 132
DB 58 IAQVAALAAADNAVAVELAETAPAGVKASSDWMESMKNKTDSYGVTTGGFATSHRTKQ 117
QY 133 AVSLQKALIEHOCAGVPTXSFSVSGRGLENTLPLEVVRGAMVRYNSLGRSHAVATV 192
DB 118 GGLQKELIFELNAGT-----FGNGTESNFALEPHSATRAAMLVRIITLLQYSGIRFE 170
QY 193 VLEALTNFNRITPVIPLRGISASAGDLSPLSYIAGAITGHPDVKVFHLHGTEKIMFA 252
DB 171 ILBALTFELNHNITPCLPRGTITASGDVLPYSIAGILTGSPNSKA-VGPNNG--EALNA 227
QY 253 REAISI.FGLEAVV-LCPKKGILGVNCTASASAMATLSLHDSMLSLSQALTALTVEMV 311
DB 228 GEAFSLAGVSGFELPKKGLALVNGTAVSGSLASIALYDANILVALISEVTSVFAEVMN 287
QY 312 GQGSFAPFIHDVCRPHGQVEVARNIRITLSSGFAYHEHEEVKVDDEGILRQDRYPL 372
DB 288 GKR-EFTDHLTKLKHHPQIEAALIMEHILDGSAY-VKAAQKHEITPLPKPQDRYAL 345
QY 372 RISPQFLGLVEDMMAAYSTLSLENTTTNDPLLVENKQTAHGNFOASAVSISMEXTR 431
DB 346 RTSPQWLGPIEVIRATKQIEREINSVNDPLDIVSRKALHGNFQGTPIGVSMNTR 405
QY 432 LALALIGKLNFTQCTELNLAAMNRLPSCIA-AEDPSLNYHGKGLDIIHAYASELGHIA 490
DB 406 LAIAAIGKLMFAQFSLVNDFFYNNGLPSNLSSGRNPISLDYGRKGEIIMAAVCEYELQFLA 465
QY 491 NPVTTFVQPAKMGNOAVNSLALISARTRAEADVLSLLASHLYCTTQAVDLAMELDFK 550
DB 466 NPVTNHVQSAECHNQVNSLIGLISRKTEALDIKLMSSTYLIALCQAVDLRHLDE 521
QY 551 KQDPLPLPTLLQGHSTGLDVNALALEVKKALKRLEQTTTYDLEPRHDA----- 601
DB 522 -----ENLRISVKNVTSQVAKKILTLNGLNGELHPSRFCEKLLR 560
QY 602 ---FSY---ATGT-----VVELLSSSPS-ANVTLTAVNAKVASAKAI 638
DB 561 VVDREYVFAIYIDPCGTYPLMGKRLQVLDHALNNGSEKNVTSISIQKIEAREVELKA 620
QY 639 SLTRFVRN-RFWGTBSSQAPALAYLSPTRVLYSVREELGVQARGGVFVQOQETIGS 697
DB 621 ILRKEVESARIALBESGNPAIGNRTECRSYPLKFRIRBELGTNYLTGKVV-----SPGE 675
QY 698 NVSRIYEAIKGRINFEVLVKKL 719
DB 676 ECDKVFALSKGLVDPDLKCL 697

RESULT 8
Q85251
ID Q85251 PRELIMINARY; PRT; 715 AA.
AC Q85251;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN GPCPAL3.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulids; Apiales; Apiaceae; Apioidae; Scandiceae; Daucinae;
OC Daucus.
OC NCBI_TaxId=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Kurodagosun;
RA Ozeki Y., Chikagawa Y., Kimura S., Soh H., Maeda K., Pornsiriwong W.,
RA Kato M., Akimoto H., Oyanagi M., Fukuda T., Koda T., Itoh Y.,
RA Yamada A., Ueno H., Takeda J.;
RT "Putative cis-elements in the promoter region of phenylalanine
RT ammonia-lyase gene of carrot involved in anthocyanin biosynthetic
RT pathway";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB069813; BAC56977.1; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia ligase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0005559; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L-Asparagine-lyase.
DR InterPro: IPR001106; Phe/His NHase.
DR InterPro: IPR005922; Phe am lyase.
DR Pfam: PF00221; PAL_1.
DR TIGRfams: TIGR01226; phe.am.lyase; 1.
DR PROSITE: PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SEQUENCE 715 AA; 77802 MW; 8D067795D57565B4 CRC64;

Query Match 28.1%; Score 1013; DB 10; Length 715;
Best Local Similarity 36.1%; Pred. No. 2.6e-59;

Matches 265; Conservative 131; Mismatches 275; Indels 64; Gaps 18;
QY 16 FTNGSHAAPTSAAGPTSLRRTP-----GLDGHAAHQSLEIVQELISDPTDVELSGY 71
DB 3 YTNHGH-----ENGNGVVDLCMKEDPLSGVAAEALKSHHEVRYMAEFKRPVYVTKGG 58
QY 72 SLTVRDVVG-AARKGRVRYONDEIRARYDKSVDFLKQLONSV--YGVTTGGGSADT 128
DB 59 TLISQVAALISARDDSGVKVLESAABAQKASSDWMESMKNKTDSYGVTTGGFATSHR 118
QY 129 RTEDAVSLQKALIEHOCGVPTXSFSVSGGLE--NTLPLEVVRGAMVRYNSLITRG 185
DB 119 RTKQGALQKELIFELNAGI-----FGSGAEAGNNTLPHSATRAAMLVRIITLLQ 169
QY 166 HSAVAVLVLEALTNFNRITPVIPLRGISASAGDLSPLSYIAGAITGHPDVKVFHLHG 245
DB 170 YSGIRFELIETFLNHNITPCLPRGTITASGDVLPYSIAGILTGSPNSKA---VGP 226
QY 246 TEKIFASEAALSLELEA--VLGPKKGLGVNCTASASAMATLSLHDSMLSLSQALT 303
DB 227 TGVTLSPPEARFKLAGVBSGFELPKKGLALVNGTAVSGMAVLEFAMILVLEVMVS 286
QY 304 ALTVFAMTGGQGSFAPFIHDVCRPHGQVEVARNIRITLSSGFAYHEHEEVKVDDEG 363
DB 287 ALFAEVMQGP-EFTDHLTKLKHHPQIEAALIMEHILDGSAY-VKAAQKHEMDPLK 344
QY 364 LEODRYPLRTPQFLGLVEDMMAAYSTLSLENTTTNDPLLVENKQTAHGNFOASAV 423
DB 345 PKQDRYALRTSPQWLGPIEVIRATKQIEREINSVNDPLDIVSRKALHGNFQGTPI 404
QY 424 SISKEKTLALALIGKLNFTQCTELNLAAMNRLPSCIA-AEDPSLNYHGKGLDIIHAY 482
DB 405 GVSXMDNTLALIAAIGKLMFAQFSLVNDFFYNNGLPSNLSSGRNPISLDYGRKGEIIMAA 464
QY 483 ASEIGHLANPVTTFVQPAKMGNOAVNSLALISARTRAEADVLSLLASHLYCTTQAVDL 542

```

Db      465 CSELOFLANPVNHNQSAEHNQVNSIGLISRKTSBEAVILKMTSTFLVNGICQALD 524
QY      543 RAMELDFKKQEPDPLPTLLQOHLGTGLDVNALALEVKKALKRLEQ-----568
Db      525 RHLEENLKSIVANVTQVAKRYLTWG--VNG-ELHPSRFCEKDLRAVDREYIFAYIDP 581
QY      589 -TTTTDLERHMDASVATGTVVELLSSPSANVTLTVANAMKVASAKAIS--LTREVR 645
Db      582 GSATTPLNQKRE-----TLVEHALNNGDKERNLSTISFOKIAFEDKALKPREVE 634
QY      646 N-RFWQTSQAPAHAYLSPRTVLVSFVREELGVQARRGDVFGVQOETIGSNVSIYE 704
Db      635 SARAIVESGNPAIPRIKRCRSYPLKFFVREELGT-----EYLGEKVTSPGEEDKVF 689
QY      705 AIKDGRIINVLVKML 719
Db      690 AMTKGELIDPLECL 704

RESULT 9
ID      08H6V6      PRELIMINARY;      PRT;      714 AA.
AC      08H6V6;
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      Phenylalanine ammonia-lyase.
GN      PAL1.
OS      Populus tremuloides (Quaking aspen).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosid1; Malpighiales; Salicaceae; Salicaceae; Populus.
OX      NCBI_TaxID=3693;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22264006; PubMed=12376645;
RA      Kao Y.Y., Harding S.A., Tsai C.J.;
RT      "Differential expression of two distinct phenylalanine ammonia-lyase
RT      genes in condensed tannin-accumulating and lignifying cells of quaking
RT      aspen.";
RL      Plant Physiol. 130:796-807(2002).
DR      EMBL; AF480619; FAN52279.1; -.
DR      GO; GO:0005737; Cytoplasm; IEA.
DR      GO; GO:0016211; F:ammonia ligase activity; IEA.
DR      GO; GO:0016841; F:ammonia-lyase activity; IEA.
DR      GO; GO:0016829; F:lyase activity; IEA.
DR      GO; GO:0003058; P:biosynthesis; IEA.
DR      GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR      InterPro; IPR008948; L-asparatase-like.
DR      InterPro; IPR001106; Phe_His_NH3lyase.
DR      InterPro; IPR005922; Phe_am_lyase.
DR      Pfam; PF00221; PAL; 1.
DR      TIGRPFAM; TIGR01226; phe_am_lyase; 1.
DR      PROSITE; PS00488; PAL_HISTIDASE; 1.
KM      lyase-.
SQ      SEQUENCE 714 AA; 77556 MW; 9BDADA21CECC35D CRC64;

Query Match      28.0%; Score 1010.5; DB 10; Length 714;
Best Local Similarity 35.5%; Pred. No. 3.8e-59;
Matches 265; Conservative 135; Mismatches 269; Indels 77; Gaps 21;

QY      8 LATTLANGTNGSHAPFTSAGPTSALARRP---GLDGHAAHSGOLETYGELLSDPTDD 64
Db      1 METITXNGVQNGS-----SESLCTGRDPLSGVAAEAMKSGHLDDEVKRMVAEYRKP 51
QY      65 VVELSGYSITLVDDVVG-AARKGRRVVQNDDEIRARVDKSVDFLKAOLQNSV--YGVTTG 121
Db      52 VVWLAQGNLTIAQVASIACHDASNVVELSEBAPRVVAVASSPMWVDSKDTSDYGVTTG 111
QY      122 FGGSAATRRFEDAVSLQKALIEHQLCGVTPTSXSSSVGRGLE--NTLPLEVVRGAMVIRV 179
Db      112 FGATSHRRTKQGGALQKELIRFLNNGI-----FNGGETCTHTLPHSATRAAMLVRI 162

```

```

QY      180 NSLRGSAVRVLVYLEALTNFNRHINTPIVPLRGSIASGDLSPISYIAGATGHPDVK 239
Db      163 NTLQGISGIRFELILSAITKLNNTITPCLPDKCTIASGDVPLSTIAGLIGSPSKA 222
QY      240 HVLHETIE-KIMFARERATISFGLBA--VLIQKEGLVNGTAVSASMATLSLHDSHLS 296
Db      223 ----TGPGVEVLDAVEAFKAAGIDSGFEELQPKEGALVNGTAVGSLAMVLFETVLA 278
QY      297 LLSQALTLTLEAVNGOQGSFAPFIHDVCRPHQGVAVANINITLSSGSFAYE---HE 352
Db      279 VLSELISATFPAEVNNGRP-EFTDHLTKLKHPEQLPAAIMEHITLDGSAVMRAAKLHE 337
QY      353 EEVKVQDEGLRQDRVYLPRTSPQFLGPLVEDMMAVSTLSJENNTTNDPLDVENKQT 412
Db      338 MDPLQKP-----KQDRVALRTPSPQWLGPGQIEVIRFSKISEREINSVNDPPLIDVSRKA 392
QY      413 AHGNGFQASAVSISNEKRLALIGKLNFTQCELLNAAMNGLPSC-LAADDPLNTH 471
Db      393 LHGNFGQTPIGVSNVYKLAIASIGLLPAQFELVNDPYNNGLPENLPAASNPLSDYG 452
QY      472 GKGLDHTAAVASELGHLANPVTTFVQPAEMGNQAVNSLALISARRTAANDVLSLLAS 531
Db      453 FKGAIAMASYCSLQYLANPVTHVQSAEHNQDVNSGLISRKTAEAVDILKMTST 512
QY      532 HLVTQLQAVDIRAMELDFKKQFDPLPTLLQOHLGTGLDVNALALEVKKALKRLEQTT 591
Db      513 FLVALCQALDIRHEENLRSAVKNVTSVSKRYLTTANB--LHPSRCEKELKVV- 568
QY      592 YDLERHMDASVATG-----TYVELSSPSANVTLTVANAMKVASAK 636
Db      569 -DRE---DVRAVMDPESATYPLMQKRLQVLYHALNGENENKASTS-FQKIAFED 622
QY      637 AIS--LTREVRN-RFWQTSQAPAHAYLSPRTVLVSFVREELGVQARRGDVFGVQOE 693
Db      623 ELKALLPREVESARAAYDSGNSALENKIKECRSYPLKFFVREELGT-----GLITGENVR 677
QY      694 TIGSNVSRITVPAIDGRINHLVKML 719
Db      678 SPGEEDFVFTAMCGKLIIDPLECL 703

RESULT 10
ID      07X6M8      PRELIMINARY;      PRT;      714 AA.
AC      07X6M8;
DT      01-OCT-2003 (TREMBLrel. 25, Created)
DT      01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE      OSJNB0073R02.16 OR OSJNB0061C13.5 protein).
GN      OSJNB0073R02.16 OR OSJNB0061C13.5.
OS      Oryza sativa (Rice).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC      Ehrhartoideae; Oryzaceae; Oryza.
OX      NCBI_TaxID=4530;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA      Liu Y.L., Wu J., Yu Z., Chen L., Fan D.L., Meng Q.J., Zhang L.,
RA      Lu Y.Q., Yu S.L., Liu S.H., Lu T.T., Zhang Y.D., Lu Y., Li C., Li T.,
RA      Zhang Y., Hu H., Jia P.X., Qian Y.M., Yang K., Zhou B., Chen Z.H.,
RA      Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA      Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J.Y., Yin H.F.,
RA      Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.W., Sheng H.H.,
RA      Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
RA      Gu J.L., Chen S.T., Ni L., Zhu P.H., Hong G.F.;
RA      Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AL731616; CAE05456.1; -.
DR      EMBL; AL731629; CAE05623.1; -.
SQ      SEQUENCE 714 AA; 76937 MW; C4C21FF6ED588FA CRC64;

Query Match      27.9%; Score 1007.5; DB 10; Length 714;
Best Local Similarity 36.4%; Pred. No. 6e-59;

```



```

QY 275 VNGAVASAMATLSLHDSHMSLSQALTAITVAMVGQGSFAFPIHIVCRPHGOVEV 334
D 259 VNGTAVSGMAVNLDAVNLALSLVSAITPAEVMGKP-EFTDLTHKIKHHPGOIEA 317
QY 335 ARNIRLLSGSSFAVEHEEVKVDDEGLRQDRYPLRTSPQFLPVEDMMAHAYSTLSL 394
D 318 AAIWEYTLDSGY-VKAAQKVHMDPLQKPKODRYALRTSPQFLPVEDMMAHAYSTLSL 376
QY 395 ENNTTNDPLDVENKQTAHGNFQASVISMETRLALALIGLANTQCTELLNMA 454
D 377 EINSVNDNPLIDVSRNKAHLGNGPQSTPIGSMNDTRLALAIIGLMAQFSSELVNFPY 436
QY 455 RGLPSCIA-AEDPSLYNHGKGLDIIHAAVASSELGLANPVTTFVQPAEMGQAVNSALI 513
D 437 NGPSPNLGGSRNPSLDYFGKGAELAAVASCSELQFLANPVTNHVQSAQHNDVNSGLI 496
QY 514 SARTAEANDVLSLASHYCTLOAVDLPAVELDFKQFDBLFTLLQHLGTGLDVNA 573
D 497 SAKRTAESEVILKLMSTYLAVALCOSIDRLHLEENIKSTVXNTSLVAKKILTTG-VNG 554
QY 574 LALEVKKALKRLEO-----TTTYDLPRMHDASFATGTVVELLSSPS 618
D 555 -ELHPSRPFCKDLRYVDEYFATYDACSATYFLMOKLQVI-----VDHALNEND 607
QY 619 ANVTLANVAMKVASAEKAI--LTREVN-RFWQTPSSQAPAHAYSPRTVLSFVRE 675
D 608 AGTSIFQ---KISEFEELKAVLPKEVEGVASAVESSTLTIPNRIKESRSPLYRVRRE 663
QY 676 ELGVQARQGVFVGVQGETIGSNVSRIVEAKIDGRI 711
D 664 ELST-----GLTGEVTSFGEEFQKVFALCKGHI 694

RESULT 14
Q8VXG7 PRELIMINARY; PRT; 703 AA.
AC 08VXG7;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC OC/CAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Corso;
RX MEDLINE=9716112;
RA Rosier J., Krekel F., Amrhein N., Schmid U.;
RT "Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase
RT activity.";
RL Plant Physiol. 113:175-179(1997).
DR EMBL; L77912; AAL40137.1; -.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; P:ammonia ligase activity; IEA.
DR GO; GO:0016829; P:lyase activity; IEA.
DR GO; GO:0045548; P:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Asparagine-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; Pal; 1.
DR TIGRfam; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SQ SEQUENCE 703 AA; 74926 MW; 11374FD68516971E CRC64;
Query Match 27.6%; Score 997.5; DB 10; Length 703;
Best Local Similarity 35.6%; Pred No. 2,8e-58;
Matches 262; Conservative 133; Mismatches 247; Indels 93; Gaps 21;

```

```

QY 29 AGPTALKRTGDLQDHAAHQ---SQLEIVQELLSDPTDDVVELSGVSLTVRDY--VGAAR 63
D 2 AANGAIVESDPLNMWAAAAELAGSHLDEVKRMVAQAPQVVKLEGSTLRGVQAAVASAK 61
QY 84 KGRVRVQNDDEIRAVYDKSVDFLKAQLON--SVYGTTFGSGADPTREDAVSLKALI 141
D 62 DASGVAVELDEARPRVYASSEWILDCIANGGDIYGVTTFGGSHRRTDQALQVELL 121
QY 142 EHQLCGVPTXSSESVKRGENTLPLEVVRGAMVIVNSLTHGKAVRLVIEALTNPL 201
D 122 RHVANGIGTSDG-----HTLPSEVRAALVAINTLLQYSGIREILIEATTKLL 173
QY 202 NHRITPIVPLRGSISASGDSLPLXYIGAITGHPDVVYHLHGETEKIMAREARISLFG 261
D 174 NTGVSPCLPLKGTITASGDVPLSLYIGLITGRNQA-VTVGGRK--VDMAAFKLAGI 230
QY 262 EA-VVIGPKKGLVNGTAVASAMATLSLHDSHMSLSQALTAITVAMVGQGSFAP 319
D 231 EGGFKLNPKKGLAI VNGTSVGSALAAATVMYDAVLAIVSEVSAVCEVMNGKP-EYTD 289
QY 320 FIHDVCRPHGOVAVANIRTLSSGSSFAVEHEEVVKDDEGL--RODRYPLRTSPQF 377
D 290 HTHKIKHHPPOSIEAIMEHILDGSSFM--KQAKVNELDPLRKQDRYALRTSPQ 346
QY 378 LQPLVEDMMAHAYSTLSLNNNTTNDPLDVENKQTAHGNFQASVISMETRLALALI 437
D 347 LGPQIEVIRAATKSIEREVNSVNDNPVIVDHQKALGNGFQSTPIGSMNARLAIAMI 406
QY 438 GKLNFQCTELLNMAHGLPSCIA-AEDPSLYNHGKGLDIIHAAVASELGLANPVTTF 496
D 407 GKLMPAQSESLVNEPYNGLTSNLAGSRNSLDGFGGTIAVASTSELYQVGNPLTNH 466
QY 497 VQPAEMGQAVNSALISARTAEANDVLSLASHYCTLOAVDLPAVELDFKQFDBL 556
D 467 VQSANDEHNDVNSLGVLSARKTAAEIDILKMSSTYVALCOAVDLHLEENIKAS---- 522
QY 557 LPTLLQHLGTGLDVNALALEVKKALKRLEQTTTYDLERPMWDASATGTVVELLSS 616
D 523 VKNVTQVAKKVLTMNPSG-ELSGRPSKEXELSAIDREX---VTYYA-----E 567
QY 617 PSANVTLANVAMKVASAEKAI--LTREVN-RFWQTPSSQAPAHAYSPRTVLSFVRE 675
D 568 DAAASLSLMLKTLAVVDHALSGERGAGALRYLDHQRG---GAPRAAGGGGRPR 624
QY 658 -----AHAYSPRTVLSFVREELGVQARQD-----VFGVQQTITGS 697
D 625 GVAEGTAVANRIADSRSPFLYRFVREELGVFLTGERLKSPEGECNKVFVGISQ---GK 681
QY 698 NVSRIVEAK--DGR 710
D 662 LVDPMLCKIKWDGK 696

RESULT 15
Q9ATN7 PRELIMINARY; PRT; 716 AA.
AC 09ATN7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase.
GN PAL.
OS Agastache rugosa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC OC/Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC OC/Lamiales; Lamiales; Nepentaceae; Mentaceae; Agastache.
OX NCBI_TaxID=39271;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim J.-B., Kim U.-B., Cho K.-J., Moon U., Ha S.-H.;
RT "Molecular Cloning and Characterization of Phenylalanine Ammonia-
RT lyase (PAL) Related to the Biosynthesis of Rosmarinic acid in Agastache

```


(without alignments)
7040.653 million cell updates/sec

Sequence: 1 atgccccctcgtgcattc.....nnnnnnnnnnnnnnnnnnn 2475

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Maximum DB seq length: 20000000000

Post-processing:	Minimum Match 0%
	Maximum Match 100%

Listing first 45 summaries

Database ::

1:	gb	ba:	*
2:	gb	hg:	*
3:	gb	in:	*
4:	gd	om:	*
5:	gd	ov:	*
6:	gb	pac:	*
7:	gb	pat:	*
8:	gb	pl:	*
9:	gb	pr:	*
10:	gd	ro:	*
11:	gd	sts:	*
12:	gb	sy:	*
13:	gb	un:	*
14:	gb	vi:	*
15:	em	ba:	*
16:	em	fun:	*
17:	em	hum:	*
18:	em	in:	*
19:	em	ma:	*
20:	em	om:	*
21:	em	or:	*
22:	em	ov:	*
23:	em	pat:	*
24:	em	ph:	*
25:	em	pl:	*
26:	em	ro:	*
27:	em	sts:	*
28:	em	un:	*
29:	em	vi:	*
30:	em	hg	hum:
31:	em	hg	inv:
32:	em	hg	other:
33:	em	hg	mus:
34:	em	hg	pln:
35:	em	hg	rod:
36:	em	hg	mam:
37:	em	hg	vrt:
38:	em	sy:	*
39:	em	hgo	hum:
40:	em	hgo	mus:
41:	em	hgo	other:

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	218	86.0	2475	6	AR199606	AR199606 Sequence
2	2128	86.0	2475	6	AX366868	AX366868 Sequence
3	1699.4	68.7	2439	6	AR199605	AR199605 Sequence
4	1699.4	68.7	2439	6	AX366866	AX366866 Sequence
5	1677.4	67.8	2331	6	E01783	E01783 DNA sequenc
6	1677.4	67.8	2331	6	E01784	E01784 DNA sequenc
7	1677.4	67.8	2331	6	E01785	E01785 DNA sequenc
8	1676.4	67.7	2315	6	RTPAAGNA	X12702 R. toruloid
9	1629.4	65.8	2419	6	AR199602	AR199602 Sequence
10	1629	65.8	2419	6	AX366860	AX366860 Sequence
11	1626.4	65.7	2151	6	E01759	E01759 DNA encodin
12	1626.4	65.7	2151	6	AR206032	AR205032 Sequence
13	1626.4	65.7	2151	6	AR282635	AR282635 Sequence
14	1626.4	65.7	2151	6	AX082544	AX082544 Sequence
15	1626.4	65.7	2151	6	AX370659	AX370659 Sequence
16	1623.2	65.6	2151	6	E01543	E01543 L-phenylala
17	1623.2	65.6	2151	6	AR205031	AR205031 Sequence
18	1623.2	65.6	2151	6	AR282638	AR282638 Sequence
19	1623.2	65.6	2151	6	AX082542	AX082542 Sequence
20	1623.2	65.6	2151	6	AX370657	AX370657 Sequence
21	1610.4	65.1	2061	6	AR282658	AR282658 Sequence
22	1610.4	65.1	2061	6	AX370681	AX370681 Sequence
23	1598.6	64.6	2787	6	AR199604	AR199604 Sequence
24	1598.6	64.6	2787	6	AX366864	AX366864 Sequence
25	1258	50.8	3340	8	RTPA22	X51513 Rhodospirid
26	1213.6	49.0	2986	8	RHDP2L	M8261 R. toruloid
27	1156.6	49.7	3586	8	RRPAL	X13094 Rhodotorula
28	1124.4	48.4	2741	6	AX366876	AX366876 Sequence
29	450.2	18.7	2488	8	MZEP2L	L77912 Zee maya ph
30	450.2	18.2	2504	8	AK102817	AK102817 Oryza sat
31	447	18.1	2280	6	AX654420	AX654420 Sequence
32	442.2	17.3	2124	6	AX653389	AX653389 Sequence
33	442.2	17.9	2124	6	AX660330	AX660330 Sequence
34	442.2	17.9	2510	8	AK100346	AK100346 Oryza sat
35	440.4	17.8	2016	8	AY224546	AY224546 Oryza sat
36	440	17.8	2282	6	AX660254	AX660254 Sequence
37	436.8	17.6	2126	6	AX660228	AX660228 Sequence
38	429.6	17.4	138226	8	OSJUN024	AL731629 Oryza sat
39	429.6	17.4	138488	8	OSJUN0282	AL731616 Oryza sat
40	426.4	17.2	149553	8	AP005000	AP005000 Oryza sat
41	426.4	17.2	154494	2	AP005070	AP005070 Oryza sat
42	426.4	17.2	165957	8	AP006523	AP006523 Oryza sat
43	422.6	17.1	170186	2	AC135419	AC135419 Oryza sat
44	422.6	17.1	173374	8	AC135449	AC135449 Oryza sat
45	421.8	17.0	4412	8	OSPAL	X16099 Rice gene f

RESULT 1	ARI199606	2475 bp	DNA	linear	PAT 20-APR-2002
LOCUS	ARI199606				
DEFINITION	Sequence 20 from patent US 6355468.				
ACCESSION	ARI199606				
VERSION	ARI199606.1	GI:20249680			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 2475)				
AUTHORS	Yoshida, R.K. and Koetsura, A.B.				
TITLE	Phenylalanine ammonia lyase polypeptide and polynucleotide sequences and methods of obtaining and using same				
JOURNAL	Patent: US 6355468-A-20 12-MAR-2002;				

FEATURES
source Location/Qualifiers
1..2475
/organism="unknown"
/mol_type="unassigned DNA"

Query Match 86.0%; Score 2128; DB 6; Length 2475;
Best Local Similarity 100.0%; Pred. No. 7.4e-224;
Matches 2408; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

1 ATGGCCCCCTCCCTCGACTCGATCGGACCTGCTGCTCCCAAGGCTGCTCAGGATG 60
1 ATGGCCCCCTCCCTCGACTCGATCGGACCTGCTGCTCCCAAGGCTGCTCAGGATG 60
61 CAGCGCGNNCCGNNNGNCGNACGAGGCGCAGCTCCACNCTCGNCGCGCGCG 120
61 CAGCGCGNNCCGNNNGNCGNACGAGGCGCAGCTCCACNCTCGNCGCGCGCG 120
121 GCTGCTCTCTCCGACGACCCGAGNNGACGAGCTGACATGCTNGAGNAGATCCTCGCG 180
121 GCTGCTCTCTCCGACGACCCGAGNNGACGAGCTGACATGCTNGAGNAGATCCTCGCG 180
181 ACCCCACGNNACGNNCGNNNTGAACTGACGAGGATCACCTCACTCGAGNAGCTG 240
181 ACCCCACGNNACGNNCGNNNTGAACTGACGAGGATCACCTCACTCGAGNAGCTG 240
241 TCAGCGCGCGCGGAAAGGCGCGNCGTCCGCTGTCACAGNCGACGAGATCCGCGCA 300
241 TCAGCGCGCGCGGAAAGGCGCGNCGTCCGCTGTCACAGNCGACGAGATCCGCGCA 300
301 AAGATGACAAAGNNGTCAGTTCCTCGGNNCGAGCTGACAGAGGTCGAGATCCGAG 360
301 AAGATGACAAAGNNGTCAGTTCCTCGGNNCGAGCTGACAGAGGTCGAGATCCGAG 360
361 ACGATGATTTTCGCGGAGCTTCGCGCGACACCCGAGACTGAGATGCAATCTCGTCG 420
361 ACGATGATTTTCGCGGAGCTTCGCGCGACACCCGAGACTGAGATGCAATCTCGTCG 420
421 GGNCTCTCTGAGACACGAGCTTCGCGGATGTCCTCCNACGTCGATGATCTCTCGCTC 480
421 GGNCTCTCTGAGACACGAGCTTCGCGGATGTCCTCCNACGTCGATGATCTCTCGCTC 480
481 GAGCGCGGCTCGAGAACTCGCTTCGCTCGAGTGTGTCGCGCGCGCATGACATCCGC 540
481 GAGCGCGGCTCGAGAACTCGCTTCGCTCGAGTGTGTCGCGCGCGCATGACATCCGC 540
541 GTCAACTCGCTCAACGCGGCGCACTCGGCGTCCGCTGCTGCTGGAAGCGCTCAC 600
541 GTCAACTCGCTCAACGCGGCGCACTCGGCGTCCGCTGCTGCTGGAAGCGCTCAC 600
601 AACTCTCTCAACGAGGATCAACCCCATGTCCTCCGCGGAGCACTTCGCGCTG 660
601 AACTCTCTCAACGAGGATCAACCCCATGTCCTCCGCGGAGCACTTCGCGCTG 660
661 GAGGACCTCTCCGCTCTCTCATGATGCGCGCGCATGACCGGTCACCGGACNNCAAG 720
661 GAGGACCTCTCCGCTCTCTCATGATGCGCGCGCATGACCGGTCACCGGACNNCAAG 720
721 GTTCAAGTNTTACAGAGGCGAGNAGATGATGCTGCGCGGAGGCGATGCGCTC 780
721 GTTCAAGTNTTACAGAGGCGAGNAGATGATGCTGCGCGGAGGCGATGCGCTC 780
781 TTNGTCTCGAGCGCGCTCTCTCATGATGCGCGCGCATGACCGGTCACCGGACNNCAAG 840
781 TTNGTCTCGAGCGCGCTCTCTCATGATGCGCGCGCATGACCGGTCACCGGACNNCAAG 840
841 GCGCTCTCGCGCTCGATGAGGAGCTCGCTTCGACGAGCAACAATCTCTCGCTCTC 900
841 GCGCTCTCGCGCTCGATGAGGAGCTCGCTTCGACGAGCAACAATCTCTCGCTCTC 900
901 TCAGAGCGCTCAAGGCTCTTACGATCGAGGCGATGTCGACGAGCGCGCTCTCTCAC 960
901 TCAGAGCGCTCAAGGCTCTTACGATCGAGGCGATGTCGACGAGCGCGCTCTCTCAC 960

961 CATTCTCTCAAGAGCTCAAGCGCGCTTCAACCCGACCCGAGATCGAGTTCGCGCAATC 1020
961 CATTCTCTCAAGAGCTCAAGCGCGCTTCAACCCGACCCGAGATCGAGTTCGCGCAATC 1020
1021 CGCAGCTCTCTGAGGAGGAGGAGGATTTGCGCTTCCACCAAGAGGAGGATCAAGTCAAG 1080
1021 CGCAGCTCTCTGAGGAGGAGGAGGATTTGCGCTTCCACCAAGAGGAGGATCAAGTCAAG 1080
1081 GACGAGAGGAGGATCTCTCGCAGAGCGGATCCGCTCGCAGTTCGCTCGATGAGCTC 1140
1081 GACGAGAGGAGGATCTCTCGCAGAGCGGATCCGCTCGCAGTTCGCTCGATGAGCTC 1140
1141 GAGCGGCTCTGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1200
1141 GAGCGGCTCTGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1200
1201 TCGAGAGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1260
1201 TCGAGAGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1260
1261 TTCAGAGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1320
1261 TTCAGAGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1320
1321 GGCAGGCTCACTTCAAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1380
1321 GGCAGGCTCACTTCAAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1380
1381 CATTCTCTGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1440
1381 CATTCTCTGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1440
1441 GCGCGGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1500
1441 GCGCGGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1500
1501 GAGCGGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1560
1501 GAGCGGCGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1560
1561 ACNCGGAGGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1620
1561 ACNCGGAGGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1620
1621 CAGGCGGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1680
1621 CAGGCGGAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1680
1681 NCNCGCTCTCTCAAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1740
1681 NCNCGCTCTCTCAAGGAGGATTTCAAGCGCGCGCGCTCTCTGCTGAGGCGGAG 1740
1741 CTCGAGGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1800
1741 CTCGAGGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1800
1801 CTCGAGGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1860
1801 CTCGAGGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1860
1861 NNCTCTCTGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1920
1861 NNCTCTCTGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1920
1921 CCGCGAGGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1980
1921 CCGCGAGGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 1980
1981 CGTCTCTGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 2040
1981 CGTCTCTGAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAGGAGGATTTCAAG 2040

QY	241	TCGCGCCGCGNCNGCAGAGGGCCGNCNGTCCGCGTCNCACACAGCGACGAGATCCGCGCA	300
Db	241	TCGCGCCGCGNCNGCAGAGGGCCGNCNGTCCGCGTCNCACACAGCGAGATCCGCGCA	300
QY	301	AAGATGCAAAAGAGTGCAGATTCCTCCGANNCAAGTCAACAAGAGTCAACGAGTTC	360
Db	301	AAGATGCAAAAGAGTGCAGATTCCTCCGANNCAAGTCAACAAGAGTCAACGAGTTC	360
QY	361	ACGACTGGTTTCGGCGGCTCGGCGGACACCGGACCTGAGAGATGCNATCTTCGCTCCAGAG	420
Db	361	ACGACTGGTTTCGGCGGCTCGGCGGACACCGGACCTGAGAGATGCNATCTTCGCTCCAGAG	420
QY	421	GCNCTCCTGAGACACAGACTCTGGGGGTGTCCTCCNAGAGTCGATCGATCTTCNAGCTTC	480
Db	421	GCNCTCCTGAGACACAGACTCTGGGGGTGTCCTCCNAGAGTCGATCGATCTTCNAGCTTC	480
QY	481	GGNCGGGCTTCGAGAACTCGCTTCGGTCGAGGTGTCGCGCGGCATGACCAATCCGC	540
Db	481	GGNCGGGCTTCGAGAACTCGCTTCGGTCGAGGTGTCGCGCGGCATGACCAATCCGC	540
QY	541	GTCAACTCGCTACNCGCGGCGACATCGGNCGTCCGCGCTGTCGCTCGAGGGCGCTCAC	600
Db	541	GTCAACTCGCTACNCGCGGCGACATCGGNCGTCCGCGCTGTCGCTCGAGGGCGCTCAC	600
QY	601	AACCTTCCTCACAACGAGGATCACCCCATGTCCTCCCGCGGACCAATTCGCGGTG	660
Db	601	AACCTTCCTCACAACGAGGATCACCCCATGTCCTCCCGCGGACCAATTCGCGGTG	660
QY	661	GCGGACCTCTCCCNCTCTCTAATCATGCGCGCGCATACCGGTACACCCGACANNCAG	720
Db	661	GCGGACCTCTCCCNCTCTCTAATCATGCGCGCGCATACCGGTACACCCGACANNCAG	720
QY	721	GTCACGTTNTTTCACGAGGCGANNAGAAGATCATGTNCGCGCGGACAGGCGATCCGCGTC	780
Db	721	GTCACGTTNTTTCACGAGGCGANNAGAAGATCATGTNCGCGCGGACAGGCGATCCGCGTC	780
QY	781	TTCGTCCTGAGACCCGCTGCTCTTGCGCCGAGAGGAGTTCGGTCTGTCACCGCACG	840
Db	781	TTCGTCCTGAGACCCGCTGCTCTTGCGCCGAGAGGAGTTCGGTCTGTCACCGCACG	840
QY	841	GCGGTCCGCGTCGATGGGACCCCTGCTCTGACAGGACACATGCTTCGCTCCCTC	900
Db	841	GCGGTCCGCGTCGATGGGACCCCTGCTCTGACAGGACACATGCTTCGCTCCCTC	900
QY	901	TCGACAGCGCTCAGCGCTCTTACCGTGCAGGCGATGCTCGGCGCACGCGCTGTTCCAC	960
Db	901	TCGACAGCGCTCAGCGCTCTTACCGTGCAGGCGATGCTCGGCGCACGCGCTGTTCCAC	960
QY	961	CCTTCTCTCCAAGATCAGCGGCCCTTACCCGACCAGATTCGAGGTTCGCGGCAATATC	1020
Db	961	CCTTCTCTCCAAGATCAGCGGCCCTTACCCGACCAGATTCGAGGTTCGCGGCAATATC	1020
QY	1021	CGACGCTCCTCAGGGGACGANNGTTCGCTCCACACACAGAGAGAGGTCAAGTCAAG	1080
Db	1021	CGACGCTCCTCAGGGGACGANNGTTCGCTCCACACACAGAGAGAGGTCAAGTCAAG	1080
QY	1081	GACGACGAGGGCATTCCTCCGCCAGGACCGGCTACCGCATCTCGCTCAGTGGCTC	1140
Db	1081	GACGACGAGGGCATTCCTCCGCCAGGACCGGCTACCGCATCTCGCTCAGTGGCTC	1140
QY	1141	GCGCGCTGTCAGGAGCATGATTACGCCCCACGNCATCTCTCGCTCGAGGCGGNCAG	1200
Db	1141	GCGCGCTGTCAGGAGCATGATTACGCCCCACGNCATCTCTCGCTCGAGGCGGNCAG	1200
QY	1201	TCGACGACCGACACCCGCTCATCGAGTCGAGAACAGANAGACCAACGCGGCGAAC	1260
Db	1201	TCGACGACCGACACCCGCTCATCGAGTCGAGAACAGANAGACCAACGCGGCGAAC	1260
QY	1261	TTCGAGGCGCGCTGTGCGNAAACGATGAGAGAGCTGCTCGCCTCGCCTGATC	1320
Db	1261	TTCGAGGCGCGCTGTGCGNAAACGATGAGAGAGCTGCTCGCCTCGCCTGATC	1320
QY	1321	GCGAACCTCACTTACGCGAGTCAACGAGTGTCTCAACCCGCGCATGACCGCGGCTTN	1380

Db	1321	GGGAACTCACTTCAACGAGCTCACCGAATGCTCAACCGCGGCAATGAACCGCGGCTT	1360
Qy	1381	CCNTCTGCTCGCTGACCGAAGAACCCNTCGCTCTCTATCATCTGAAGGACCTCGACATT	1440
Db	1381	CCNTCTGCTCGCTGACCGAAGAACCCNTCGCTCTCTATCATCTGAAGGACCTCGACATT	1440
Qy	1441	GGCGGCMGCMTAACTTCCGAGTCCGAGCACTTGGCCAAACCCGGTAAACACCAAGCTC	1500
Db	1441	GGCGGCMGCMTAACTTCCGAGTCCGAGCACTTGGCCAAACCCGGTAAACACCAAGCTC	1500
Qy	1501	CAGCCGGCAGATGGGCAACCAAGCCGCTCACTCGCTGCNCTCATCTCGACGCGCGC	1560
Db	1501	CAGCCGGCAGATGGGCAACCAAGCCGCTCACTCGCTGCNCTCATCTCGACGCGCGC	1560
Qy	1561	ACNCGCCGAAGCCAAAGAGTCTCTTCTCTCTCTCTCGCACCACTCTAATGCGTCTC	1620
Db	1561	ACNCGCCGAAGCCAAAGAGTCTCTTCTCTCTCTCTCGCACCACTCTAATGCGTCTC	1620
Qy	1621	CAGCCGCTGACCTCGCGCGATGGAGTTGAGTTCAAGAAGAGTTGACCCGATNTTC	1680
Db	1621	CAGCCGCTGACCTCGCGCGATGGAGTTGAGTTCAAGAAGAGTTGACCCGATNTTC	1680
Qy	1681	NCNCGCTCNTCAGCAGCACTTTGGCNCNCGCCCTCGACGGCINNNTNNNNNNNACAA	1740
Db	1681	NCNCGCTCNTCAGCAGCACTTTGGCNCNCGCCCTCGACGGCINNNTNNNNNNNACAA	1740
Qy	1741	CTCGNCGACAAGGTAAACAAGCGCTCAACAAGGNTCGAGAGACCACTCGTAGAC	1800
Db	1741	CTCGNCGACAAGGTAAACAAGCGCTCAACAAGGNTCGAGAGACCACTCGTAGAC	1800
Qy	1801	CTCGAGCCGCGTGGCACGAGACCTTCTGTTGCGACACGGCACCCGCTCGACGNNNNN	1860
Db	1801	CTCGAGCCGCGTGGCACGAGACCTTCTGTTGCGACACGGCACCCGCTCGACGNNNNN	1860
Qy	1861	NNGTCTCGCCNNNNGCCANNAGTCTCGCTCGCNCGCTGCAACGCTCTGAAAGTCTGCT	1920
Db	1861	NNGTCTCGCCNNNNGCCANNAGTCTCGCTCGCNCGCTGCAACGCTCTGAAAGTCTGCT	1920
Qy	1921	CCGCCAGAAAGCATTCTGCTCAACGGCANGTCCGACNCTTCTGAGCGNCCTG	1980
Db	1921	CCGCCAGAAAGCATTCTGCTCAACGGCANGTCCGACNCTTCTGAGCGNCCTG	1980
Qy	1981	CGTGCTGCTGCGCCCGCGCTCNCGTACTCTCGCGGCAAGCGGCTCTGTACTCTGTTG	2040
Db	1981	CGTGCTGCTGCGCCCGCGCTCNCGTACTCTCGCGGCAAGCGGCTCTGTACTCTGTTG	2040
Qy	2041	TCCGGAGGAGTCCGGCGTCAAGGACCGCGCGGCGACGCTTCTCGCGGAAAGAGAG	2100
Db	2041	TCCGGAGGAGTCCGGCGTCAAGGACCGCGCGGCGACGCTTCTCGCGGAAAGAGAG	2100
Qy	2101	TGACATCGGCACCAACGCTCTCCGCACTCAAGAGGCACTCAAGNCGCGCATCAAC	2160
Db	2101	TGACATCGGCACCAACGCTCTCCGCACTCAAGAGGCACTCAAGNCGCGCATCAAC	2160
Qy	2161	AGTCTCTGCTCAAGTCTCGCANTAGNNCNCNNTCTGCTGCTNNNNCCNCCNC	2220
Db	2161	AGTCTCTGCTCAAGTCTCGCANTAGNNCNCNNTCTGCTGCTNNNNCCNCCNC	2220
Qy	2221	NNNCNNNNNNCTTNGANNCTGANTCAGNTCAGTGCNNNNCCGANNNNCCNCCNNNN	2280
Db	2221	NNNCNNNNNNCTTNGANNCTGANTCAGNTCAGTGCNNNNCCGANNNNCCNCCNNNN	2280
Qy	2281	TNNNCTNTTCTNNCTCNCNNNNAACNAGTCTNNNNCTNNNGNTNNNNNNCCNC	2340
Db	2281	TNNNCTNTTCTNNCTCNCNNNNAACNAGTCTNNNNCTNNNGNTNNNNNNCCNC	2340
Qy	2341	NGTNNNCANNNAACNTTNNNNNNANCGANNGANTTANAGNTTNCNGNCCNNNN	2400
Db	2341	NGTNNNCANNNAACNTTNNNNNNANCGANNGANTTANAGNTTNCNGNCCNNNN	2400
Qy	2401	NNNNAAAA 2408	
Db	2401	NNNNAAAA	

DB	2401	NNNNAAA	2406
RESULT 3	ARI99605	2439 bp	DNA
LOCUS	Sequence 18 from patent US 6355468.		
DEFINITION	ARI99605		
ACCESSION	ARI99605		
VERSION	ARI99605.1	GI:20249679	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2439)		
TITLE	Yoshida, R.K. and Kooststra, A.B.		
JOURNAL	Phenylalanine ammonia lyase polypeptide and polynucleotide		
FEATURES	sequences and methods of obtaining and using same		
source	Patent: US 6355468-A 18 12-MAR-2002;		
	Location/Qualifiers		
	1..2439		
ORIGIN	/organism="unknown"		
	/mol_type="unassigned DNA"		
Query Match	68.7%; Score 1699.4; DB 6; Length 2439;		
Best Local Similarity	79.4%; Pred. No. 5.1e-177;		
Matches 1913; Conservative	0; Mismatches 459; Indels 36; Gaps 6;		
QY	1	ATGGCCCCCTC	NTGACTGCATCGCAGCTCTGTCGCCCAAGCGTCTCNAAGGANTG 60
DB	1	ATGGACCCCTCG	CTGACTGCATCTCGCACTCGTTCGCAACGGCGTGCATTCGCAAG 60
QY	61	CAGCGCGNNCCG	CNNCNCNCNACGAGCGCACGCTCCACNCTGNCNCGCGCGNG 120
DB	61	CAGGCTGT-----	CAATGGGCGCTCGACCAACTC-----GGAGTCGAG 100
QY	121	GCTGCTCTCTCC	CGACCCAGANNACGCACTCGACATCGTNGAGNAGATCTTCGCCG 180
DB	101	GCTCGACCTCG	CCCAACCCAGGTCAGACATCGTCGAGAAAGATCGTCGCCG 160
QY	181	ACCCACCGGNN	ACGACGNNTCGAACCTCGACGCGGATACACCCCTCACCTCGAGACGTG 240
DB	161	CGCCGACGACT	CGA---CGCTCGAATCGACGCTACTCTGCTCAACTCTCGAGACGTG 217
QY	241	TCGCGCGCGNC	CGAAGGGCGCNCNCTCCGCTGCACGACAGTCAGATTCGCGCA 300
DB	218	TCTCGCGCGG	AGAGGGCAGGCGCTGTCCGCTCAAGAGCAG-CGACGAGATCCGCTCA 276
QY	301	AAGATCGAANA	AGNGTGAATTCCTCCGNNCAGCTGCACAAACAAGNCTTACGAGNCTC 360
DB	277	AAGATTGAACA	ATCGGTGAATTTCTTGCGCTCGCACTCTTCATGAGCTCTTACGGCGTC 336
QY	361	ACGACTGATTT	CGCGGGCTCGGCCGACACCCGAGCTAGAGATGNCATTCGCTCCGAGAG 420
DB	337	ACGACTGGA	TTTGGCGGATTCGAGACACCCGACCGAGAGGACCATTCGCTCCAGAG 396
QY	421	GCNCTCTTG	AGACCACTGTGGCGGTGTCCTCCNACGTCGATGCATTCCTTCNGCCTC 480
DB	397	GCCTCTCTG	AGACCACTGTGGCGGTTCCTCCCTTCGTCGACATCGTTTCGCGCTC 456
QY	481	GGCGCGGCG	CGCTGAGAACTCGCTTCGCGCTGAGAGTGTCTCGCGCGCCATCGACATTCGC 540
DB	457	GGCGCGGCT	CTGAGAACTCGCTTCCTCCCTGGAAGTTGTTCGGCGGCATGACATTCGC 516
QY	541	GTCACTGCT	CANCGCGCGCACTCGGCGNCTCCGCTGTGCTCTCTCGAGGCGCTCAC 600
DB	517	GTCAACAG	CTTGACCCCGCGGCACTCGGCGTTCGCTGTGCTCTCGAGGCGCTCAC 576
QY	601	AACCTTC	CAACAAGGATCACCGCATGTCGCCCTCCGCGGACCATTCGCGCGTCG 660
DB	577	AACCTTC	CAACAAGGATCACCCCATGTGCTCCCTCCGCGGACCATTCGTGCGTCG 636
QY	661	GCGGAC	CTTCGCCNCTCCTCAATGCGCGCGCATACCGGTACACCGGACNNCAAG 720


```
/db_xref="GI:18698156"
/db_xref="RENTREMBL:CAD23831"
translation="MAPSLDLSHSFPAKVASAKOAVNGASTNLAVGSHLPTTQVTO
VDIVERKMLAAPDSTLSDGLYSLNLDGVASARKR PVRVKDSDEIRSKIDKVEFLR
SOLSMYVGTTFGGADRTEDIASLOKALLEHQLGVLPSSEFRLGRJENSL
PLEVRGAMTIRVNSLTRGHSARVLVLEALTNPLNGITPIVPLRGITSAAGLSPL
SYTAAASGHPDSKVHVHEKELIYAEEMALFMLEPVYGVKREGIGLVNGAVSA
SMATLADHMMILSLOSILMTVYAWGAGHFRPLHWHYRHPPIQIYVANTIK
LLEGRFPAVHHEEVKYKDEGILRODRIPLRTSPQWJGPLVSDILHHAFLTEAQ
STTDNPLIDVEMKTSHGNGFQAAAVANTMEXTRILGLOIGKLNFTQTEMPNMR
GLPSCIAEDPISLHYCKGLIDIAAAYISELGHANPLTTHVOPAEHMAQVNSLAI
SARRTESNDVLSILLATHLVCLQALIDRAIEFPKXFGPAIVSLIDQFGSAMG
SNRDELVEKNTLAKLEQNSYDLVPRMHDASFLAGTVEVLSLSTSLAANA
MKVAAESAIISLROVRETFMSAASSPSALSYLSPROIIYAFREBLGVKARGDV
FLQKQEVITIGSNISKIYEAIKSGRINNVILKRLA"
```

Query Match 68.7%; Score 1699.4; DB 6; Length 2439;

Best Local Similarity 79.4%; Pred. No. 5.1e-177;

Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

```
QY 1 ATGGCCCCCTCCCTCCAGCTCGATCGGCACTCGNTGCGCAAGCGNTCNMAACGAGTNG 60
Db 1 ATGGCACTCTCGCTCGACTCGATCTGCACTCGTTCGCAAGCGGCTGCACTCGCAAG 60
QY 61 CAGCGCGNNCCGNNNGNCGNACGCGGCGCAAGTCCACNCTGNCNGCGCGCGNG 120
Db 61 CAGGCTGT-----CATGCGGCTCGACCAACTC-----GCAATCGAG 100
QY 121 GCTGCTCTCTCCGACCAACCAGNNGACCACTGCAATCGTNGAGNAGTCTTCCGCG 180
Db 101 GCTCGCACTCTCGCAACCAAGTCAACCAAGTCAACATCGTCAAGAAAGTGTCTCGCG 160
QY 181 ACCCCACCGNNAAGNAGTGTCTCGCGGAGTCAACCTCACTCGAGNAGCGTNG 240
Db 161 CGCGGCGGCTCGA---CGCTGAACTCGACGCGTACTCGTCAACTCGGAAGCGTNG 217
QY 241 TCGGCGCGCGNCGCAAGGCGCGNCGTCCGCTCNAGACAGNCGAGAGTCCGCGCA 300
Db 218 TCTCGGCGCGGAGGAGGCGCGCTGTCCGCTCAAGAGACAG--GAGAGAGTCCGCTCA 276
QY 301 AAGATGACAAAGNAGTGTCTCGCGNCGNCGTCAACCAAGAGTCTTACGAGTCTC 360
Db 277 AAGATTGACAAATCGATCGATGTCTTGCGCTGCAACTCTCCATGAGGCTTACGAGGCTC 336
QY 361 ACAGCTGATTCGCGCGCTCGGCGGACACCCGAGTGAAGTGCNATCTCGCTCAAGAG 420
Db 337 ACAGCTGATTTGGCGGATCCGAGACACCGGACCGAGAGCGCATCTGCTCAAGAG 396
QY 421 GGNCTCTCGAGACCAAGCTTGGCGGTGNTCTCCNAGTGTGANTCTTCCGCTC 480
Db 397 GCTCTCTCGAGACCAAGCTTGGCGGTGNTCTCCGCTTGTGATGATCTTCCGCTC 456
QY 481 GAGCGGCGCTCGAGACTCGCTCGCTCGAGGTGCTCGGCGGCGCATACCATCGCG 540
Db 457 GCGCGGCGTCTCGAGACTCGCTTCCCTCGAGGTGTTGTCGCGGCGCATACCAATCGCG 516
QY 541 GTCACTCGCTCAACNCGCGGCACTCGGCGTTCGCGCTGTGCTTTCGAGGCGCTCAC 600
Db 517 GTCAACAGCTTGACCGCGGCGCACTCGGCTGTCCGCTGTGCTGCTGAGAGCGCTCAC 576
QY 601 AACTTCTCAACACGAGGCACTCAACCCGATCGTCCGCTCCGCGGAGCATCTCGGCGT 660
Db 577 AACTTCTCAACACGAGGCACTCAACCCGATCGTCCGCTCCGCGGAGCATCTCGGCGT 636
QY 661 GCGGAGCTTCCGCTCTCTCATACATCGCGCGCGCATCAACGATCAACCGGACNNAG 720
Db 637 GCGGAGCTCTCTCTCTCTCTACATGAGCGGCGCATCAAGGAGTCAACCGGACAGAG 696
QY 721 GTNCAAGTNNNCAAGAGGCGANNAGAAAGTCAATGTCGCGCGCGGAGGAGTCTGCGCTC 780
Db 697 GTGCAAGTGTCTCAAGAGGCAAGAGAGATCTGTACGCGCGGAGGAGTGTGCGCTC 756
```

```
QY 781 TTNGGCTCGAGCCCGCTGCTCTCGGCGCGAGAGGAGTCTCGGCTCTGCAACGCGAGC 840
Db 757 TTCAACCTCGAGCCCGCTGCTCTCGGCGCGAGAGGAGTCTCGGCTCTGCAACGCGAGC 816
QY 841 GCGGCTCTCGGCTCTGAGCGGACCGCTCGCTCTGACGACGACGACGACGCTCTGCTCTC 900
Db 817 GCGGCTCTGAGCTCATGCGGACCGCTCTGCTCTGACGACGACGACGACGCTCTGCTCTC 876
QY 901 TCGCAGGCGCTCAACGCTCTCAACGCTCGAGGCGCATGCTCGGCGCGCGCTGCTTCCAC 960
Db 877 TCGCAGTCTGCTCAACGCTCTCAACGCTCGAGGCGCATGCTCGGCGCGCGCTGCTTCCAC 936
QY 961 CCNTTCTCTCAACGCTCAACGCGCGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCATC 1020
Db 937 CCTTCTCTCAACGCTCAACGCGCGCGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCATC 996
QY 1021 CGCAGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 997 CGCAAGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
QY 1081 GACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
Db 1057 GACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
QY 1141 GCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 1117 GCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
QY 1201 TCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 1177 TCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
QY 1261 TTCAGGCGGCGGCTGTCGCAACGAGTGAAGAGTCAAGCTCGGCTCGGCTCGGCTCGGCT 1320
Db 1237 TTCAGGCGGCGGCTGTCGCAACGAGTGAAGAGTCAAGCTCGGCTCGGCTCGGCTCGGCT 1296
QY 1321 GCGAAGCTCAACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
Db 1297 GCGAAGCTCAACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
QY 1381 CGNTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
Db 1357 CCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1416
QY 1441 GCGCGGCGGAGTCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
Db 1417 GCGCGGCGGAGTCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
QY 1501 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 1477 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1536
QY 1561 AGNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Db 1537 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1596
QY 1621 CAGGCGGCTGCACTTCCGCGGAGTGAAGTTCGAGTTCGAGGAGGAGTTCGAGGAGGAGTTC 1680
Db 1597 CAGGCGGCTGCACTTCCGCGGAGTGAAGTTCGAGTTCGAGGAGGAGTTCGAGGAGGAGTTC 1656
QY 1681 NCGNCGCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 1657 GCTCTGCTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
QY 1741 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
Db 1717 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
QY 1801 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
Db 1777 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
QY 1861 NNGTCTCTGCGNNNCGCANNAAGTCTGCTGCGCGGCGGCTCAACGCTGGAAGTCTGCT 1920
```

Db 1837 TCGTGAAGTC-----GCTTCGTCTCGCCGCGCTCAACGCTTGAAGTCCCG 1885
 QY 1837 TCGTGAAGTC-----GCTTCGTCTCGCCGCGCTCAACGCTTGAAGTCCCG 1885
 QY 1921 CCGCCGAGAGGCGCATCTTCGCTCACGCCGCAAGTCCGCAACCTTCTGCAGCCGCT 1980
 Db 1886 CCGCCGAGGCGCGCATCTTCGCTCACGCCGCAAGTCCGCAACCTTCTGCAGCCGCT 1945
 QY 1981 CCGTCCGCGCGCGCGCTCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2040
 Db 1946 CCACT 2005
 QY 2041 TCCGCGAGAGGCTCGGCGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
 Db 2006 TCCGCGAGAGGCTCGGCGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2065
 QY 2101 TACAGATCGGCG 2160
 Db 2066 TACAGATCGGCG 2125
 QY 2161 ACGTCT 2220
 Db 2126 ACGTCT 2185
 QY 2221 NNNCCNN 2280
 Db 2186 TATCCCGCTGCACTCTTAGGACTGCTTCTTCTGGAATCTGCAATCTGCTTCTT 2245
 QY 2281 TNNNNCTNNCTNNCTNNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
 Db 2246 TCGTCTTCTGCT 2304
 QY 2341 NCTNN 2400
 Db 2305 AGTACCCTTCAGAGATCGATTAATCAGGAGAGATCTAGCTTGGGAGCTTCTTG 2364
 QY 2401 NNNNNAAA 2408
 Db 2365 CCGATAAA 2372

RESULT 5
 E01783 2331 bp DNA linear PAT 29-SEP-1997
 Locus E01783
 DEFINITION DNA sequence coding for L-phenylalanine ammonialyase.
 ACCESSION E01783
 VERSION E01783.1 GI:2170036
 KEYWORDS Rhodospiridium toruloides
 SOURCE Rhodospiridium toruloides
 ORGANISM Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
 Microbotrymycetidae; Sporidiobolales; Rhodospiridium.
 1 (bases 1 to 2331)
 Fukuhara,N., Yoshino,S., Matanabe,M., Nakajima,Y. and Makiguchi,N.
 RECOMBINANT PLASMID FOR MANIFESTING L-PHENYLALANINE-AMMONIA-LYASE
 AND TRANSFORMANT CONTAINING SAID PLASMID
 Patent: JP 198317086-A 1 26-DEC-1988;
 JOURNAL MITSUBI TOHATSU CHEM INC
 COMMENT OS Rhodospiridium toruloides
 PN JP 198317086-A/1
 PD 26-DEC-1988
 PF 18-JUN-1987 JP 1987152357
 PI FUKUHARA NOBUHIRO, YOSHINO SADAOKI, WATANABE MIDORI, PI
 NAKAJIMA YOSHIOYUKI,
 PI MAKIGUCHI NOBUYOSHI
 PC C12N15/00, C12N1/20//C12N9/89, (C12N1/20, C12R1/01), (C12N9/88, PC
 C12R1/01);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=JPO 559 (ATCC 10788);
 FH Location/Qualifiers

FT CDS 1..2151
 FT
 3'UTR 2152..2331.
 FEATURES
 source location/Qualifiers
 1..2331
 /organism="Rhodospiridium toruloides"
 /mol_type="genomic DNA"
 /db_xref="taxon:5286"
 ORIGIN
 Query Match 67.8%; Score 1677.4; DB 6; Length 2331;
 Best Local Similarity 80.4%; Pred. No. 1.3e-174;
 Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;
 1 ATGGCCCCCTCCNTCGACTGATCGCACTCGTCCGCAACGCGTCCGCAACGCTG 60
 Db 1 ATGGCACTCCGCTGATGATCTCGCACTCGTCCGCAACGCGTCCGCAACGCTG 60
 QY 61 CAGCGCCG 120
 Db 61 CAGGCTGT-----CAATGGCGCTTGAACAACCTC-----GCAATCGCAG 100
 QY 121 GCTGCTCTCCCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 Db 101 GCTGCACTCTGCGCAACCGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 160
 QY 181 ACCCGACCGGNNACGNNACGNNACGNNACGNNACGNNACGNNACGNNACGNNACG 240
 Db 161 CGCGCAGCAGCTCGA---CGCTCGAATCGAGCGCTACTGCTCAACTCGGAGACCTCG 217
 QY 241 TCGGCG 300
 Db 218 TCTGCG 276
 QY 301 AAGATCGACAAAGAGTGAAGTCTCTCCGNNCAGCTCNAACAAGGCTCAAGGCTC 360
 Db 277 AAGATTGAACAATCGTGAAGTCTTGGGCTCGCACTCTCATGAGAGTCAAGGCTC 336
 QY 361 AAGATCGGTTTGGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
 Db 337 ACGACTGATTTGGGCGATCGCGAGACCGCGACCGCGACCGCGACCGCGATCTCGTCAAG 396
 QY 421 GCGTCTCTGAGACACAGCTCTGCGGTGCTCCGNNACGTCGATCGATCTCTGAGCTC 480
 Db 397 GCTCTCTGAGACACAGCTCTGCGGTGCTCCGNNACGTCGATCGATCTCTGAGCTC 456
 QY 481 GCGCGCGCGCTCGAGAACTCGTTCGCTCGAGTCTGCGCGCGCGCGCGCGCGCGCG 540
 Db 457 GCGCGCGGTCTGAGAACTCGTTCGCTCGAGTCTGCGCGCGCGCGCGCGCGCGCG 516
 QY 541 GTCAACTGCTGACG 600
 Db 517 GTCAACGCTTGAACCG 576
 QY 601 AACTCTCTCAACCAAGGATCAACCGCATGTCCTCTCGCGCGCGCGCGCGCGCGCG 660
 Db 577 AACTCTCTCAACCAAGGATCAACCGCATGTCCTCTCGCGCGCGCGCGCGCGCGCG 636
 QY 661 GCGGACCTCTCCGCT 720
 Db 637 GCGGACCT 696
 QY 721 GTNCAAGTNNTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
 Db 697 GTNCAAGTNNTCAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756
 QY 781 TTNGTCTTGAAGCGGTCT 840
 Db 757 TTCAACTCGAGCGCGGTCT 816
 QY 841 GCGGCTCGGCTCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
 Db 817 GCGGCTCGGCTCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 876

QY	901	TCGACAGGCGCTCA	CGGCTCMA	CGGTGAGGGC	CA	TGATGCGCAC	CGCGGCTGGTTCAC	960
Db	877	TCGAGTTCGCTCA	CGCGCCAT	TGACGCTCGAAGGCA	TGATGGCA	CGCGGCTGGTTCAC	936	
QY	961	CCNTTCTCTCA	CGACGTCA	CGCGCCTCA	CCCGAC	CCGAGTGAAGTTCGCGCA	1021	
Db	937	CCCTTCCTTCA	CGACGTCA	CGCGCCTCA	CCCGAC	CCGAGTGAAGTTCGCGCA	996	
QY	1021	CGCA	CGCTCC	TCGAGGGC	CAGC	NNGTGTGCGGTCCAC	CGAGGAGGAGTCAAGTCAAG	1081
Db	997	CGCAAGCTCCTCGA	GGGAAGCCGCTT	TGCTGTCCACA	TGAGGAGGAGTCAAGTCAAG	1051		
QY	1081	GAGCAGCAGGAGCA	TTCTCGCCAGCA	CGGCTAC	CCCGCTCCGCA	CGTTCGCTCAGTGCCTC	1141	
Db	1057	GACGACAGAGGAGCA	TTCTCGCCAGCAGC	CGCTAC	CCCGCTTCGCA	CGTTCGCTCAGTGCCTC	1111	
QY	1141	GGCCCGCTCGTCA	CGCAGCAT	GAATTAC	CGCCCA	CGCGNCCCTTCGCTTCGAGGGCCGACAG	1201	
Db	1117	GGCCCGCTCGTCA	CGCAGCCTCAT	TACGCGCCAC	CGCGCTTCACAT	TCGAGGCGCGCACG	1171	
QY	1201	TCGACGACCGCACA	CCCGCTCAT	TCGACGTTCGAGAA	CAAGANGA	CCCA	CGCGCGCAAC	1261
Db	1177	TCGACGACCGCACA	CCCTCTCAT	TCGACGTTCGAGAA	CAAGATTG	CGACAC	CGCGCGCAT	1231
QY	1261	TTCCAGGCGAGCGCTG	TCGCGNAA	CGAGATGAGAGAA	AGTCGCGCTCGCGNCTCGCGCTGACAT	1321		
Db	1237	TTCCAGGCTCCGCTG	TCGCGCAAC	CAATGAGAGAGT	CGCTTCGCGCTTCGCGCTGACAT	1291		
QY	1321	GGCAAGCTCAACTTCA	CGCAGCTCA	CCGAGATGCTCAA	CGCGCGCATGAA	CCGCGGCTCM	1381	
Db	1297	GGCAAGCTCAACTTCA	CGCAGCTCA	CCGAGATGCTCAA	CGCGCGCATGAA	CCGCGGCTCM	1351	
QY	1381	CCNTTCTCGCTCGCTG	TCGCGAGAC	CCGCGCTTCCTAT	CACTGAGCAAGGCGCTTCGACAT	1441		
Db	1357	CCCTTCGCTCGCGCGC	CGAACA	CCCTTCGCTTCCTCA	CACTGAGCAAGGCGCTTCGACAT	1411		
QY	1441	GGCGCNGCNGCNTACA	CTTCGAGCTTCGAGCA	CGCTTCGCA	CCCGGTACGACCA	CGTTC	1501	
Db	1417	GGCGCTCGCGCTGAC	CACTTCGAGTTCGAGCA	CGCTTCGCA	CCCGGTACGACCA	CGTTC	1471	
QY	1501	CAGCGCGC	GNAGATGAGGCA	CCAGGCGCTGCA	ACTTCGCGNCTCAT	CTTCGCGNCGCGC	1561	
Db	1477	CAGCGCGCTGAGAT	TGGCGAA	CCAGGCGCTGCA	ACTTCGCGCTCAT	CTTCGCGCTGCGC	1531	
QY	1561	ACNCGCCGAGGCCAA	CGACGTCTTTCTCTCT	CCTTCGCGCA	CCCACTCTATCTGCGTCTC	1621		
Db	1537	ACGACCGAGTCCAA	CAACGCTCTTCTCTCTCT	CTTCGCGCA	CCCACTCTATCTGCGTCTC	1591		
QY	1621	CAGGCGCTGCA	CTTCGCGCGCAT	TGAGTTCGAGTTCAGAA	GAAGAGTTCGACCCGATNNTC	1681		
Db	1597	CAGGCGCTGCA	CTTCGCGCGCAT	TGAGTTCGAGTTCAGAA	GAAGAGTTCGACCCGATC	1651		
QY	1681	NCNNCGCTONT	CNAGCAGCAT	CTTTGGC	NCNGCCTTCGAC	GGCGNNNNNNNNNNNNN	1741	
Db	1657	GTCTCGCTCAT	CGACCGACAT	TTTGCGCTTCGCGCAT	TGACCGCGCGCA	CGTTCGCGCGACGAG	1711	
QY	1741	CTCGANAGACAGGTCA	CAAGCGCTTCNACAA	GAGGANTTCGAGCAGAC	CAAACTCGTAGAC	1801		
Db	1717	CTCGTAGAAGATGTA	CAAGCGCTTCGCGCA	AGGCTTCGAGCAGAC	CAAACTCGTAGAC	1771		
QY	1801	CTCGAGCGCGCTG	GGCAGCAGCGCTT	CTCGTTGCGCA	CCGCGCAC	CGTTCGTCAGANNNNN	1861	
Db	1777	CTCGTTCGCGCTG	GAGCAGCGCTTCTTCCTTCGCGCGCGCAC	CGCTTCGTCAGAGTCTCTC	1831			
QY	1861	NNGTCTCGCC	NNNNNGC	ANNAGTCTCGCTCGCGN	CGCTTCGACGCGCTTCGAGAGTTCGCT	1921		
Db	1837	TGCTGACGTC	-----GCTTCGCTCGCGCGCGTCA	CGCTTCGAGAGTTCGCT	1881			
QY	1921	CGCGCAGAGGAGCAT	CTTCGCTCA	CGCGGANNAGTTCGCGCA	CNCTTCTGAGCGC	NCCTG	1981	
Db	1886	CGCGCAGAGGAGCAT	CTTCGCTCA	CGCGGANNAGTTCGCGCA	CNCTTCTGAGCGC	NCCTG	1941	

QY	1981	CGTCGTGCTGCGCCCGGCTCNCGTACCTCTGCGCGCGAGCGGCTCTGTACTCGTGC	204
Db	1946	CGACTGTGTGCCCCGCGCTCTGTACTCTTGC CGCGACTCAAGATCTCTTACGCTTGC	200
QY	2041	TCGCGAGGAGCTCGGCGTCAAGGCGCGCGCGGAGCGTCTTCTCGGCAAGCAGAGG	210
Db	2006	TCGCCAGAGAGCTTGCGCGTCAAGCGCCGCGGAGAGGTCTTCTCGGCAAGAGAGG	206
QY	2101	TGAGGATCGGACCAACGCTTCCCGCATCTACGAGGCGATTAAAGNCGCGCATCAAC	216
Db	2066	TGAGGATCGGCTGAACGCTTCCAGATCTACGAGGCGCATTAAGTCGGGAGAGATCAAC	212
QY	2161	ACGTCCTCGTCAAGATGCTCGGCTAGNNNCCNNNNNCCNNNCTCGCTNNNNNCCNNNC	222
Db	2126	ACGTCCTCCCTCAAGATGCTCGCTGTAGACACTTCCCACTCTCGCATCCCTTCCATACC	218
QY	2221	NNNCNNNNNNNNCTTTNGNNNTGCANTCTGNCNNNNNCGAGNNTTNNCNNNNNN	228
Db	2186	TATCCCGCTCGACTTTTAGACTCGCTTCTTGTTCGACTCGATCTCGCATCGCTTCTT	224
QY	2281	TNNNCNTNNCTNNCTCNCNNNANCCNCTCNNNNNNNNNNNNNNNNNNNNNNNN	234
Db	2246	TCGTTCTTGAGCTCCCTCTC-TGACCGCGTTCGGATTATTACTCGAGATTGGAATACAGC	230
QY	2341	NGTNNNCANNA	2352
Db	2305	AGTACCATCA	2316
RESULT 6			
LOCUS	E01784	2311 bp	DNA linear
DEFINITION			PAR 29-SEP-1997
ACCESSION	E01784		
VERSION	E01784.1	GI:2170037	
KEYWORDS	JP 1988317087-A/1.		
SOURCE	unidentified		
ORGANISM	unidentified		
REFERENCE	1 (bases 1 to 2311)		
AUTHORS	Fukuhara,N., Yoshino,S., Matanabe,M., Nakajima,Y. and Makiguchi,N.		
TITLE	CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF		
JOURNAL	ADVENTITIOUS GENE PRODUCT USING SAID CONTROL		
COMMENT	Patent: JP 1988317087-A 1 26-DEC-1988; MITSUI TOATSU CHEM INC OS Rhodospiridium PN JP 1988317087-A/1 PD 26-DEC-1988 PF 18-JUN-1987 JP 1987152358 PI FUKUHARA NOBUHIRO, YOSHINO SADAO, MATANABE MIDORI, PI NAKAJIMA YOSHIYUKI, MAKIGUCHI NOBUYOSHI PC C12N15/00,C12N9/88,(C12N9/88,C12R1:01); CC strandedness: Double; CC topology: linear; CC hypothetical: No; CC anti-sense: No; CC *source: strain=IFO 559 (ATCC 10788); FH Key Location/Qualifiers FH *source: strain=IFO 559 (ATCC 10788); FH Key Location/Qualifiers FT CDS 1..2151 /product='L-phenylalanine ammoniolyase' FT FT FT 2152..2331. 3'UTR Location/Qualifiers 1..2331 /organism='unidentified' /mol_type='genomic DNA' /db_xref='taxon:32644'		
FEATURES			
SOURCE			
ORIGIN			
Query Match	67.8%	Score 1677.4;	DB 6; Length 2311;
Best Local Similarity	80.4%	Pred. No. 1.3e-174;	
Matches 1891; Conservative	0;	Mismatches 425;	Indels 36; Gaps 6;

Db 2126 ACCGCTCCCTCAAGATCTCGTTAGACACTCTTCCCACTCTCGATCCCTTCCATACC 2185
Qy 2221 NNCCNNNNNNNNCTTNGNNNTCGNNTCTGNCNNNNCGANNNTNNNNNNNN 2280
Db 2186 TATCCCGCTGCACTTAGAGACTCGCTTCTTGTCGACTCGGACTCGCTTCTT 2245
Qy 2281 TNNNNCTNNCTNNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
Db 2246 TCGTTCTTGCTGCTCTCTC-TAGACCGTGTGGTATTACTCGAGATTGTGAATCAAGC 2304
Qy 2341 NGTNNNCANNA 2352
Db 2305 AGTACCATCCA 2316

RESULT 7
E01785
LOCUS E01785 2331 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA sequence coding for L-phenylalanine ammonia-lyase.
ACCESSION E01785
VERSION E01785.1 GI:2170038
KEYWORDS JP 1988317088-A/1.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.

REFERENCE
AUTHORS Nakajima, N., Yoshino, S., Watanabe, M., Yamamoto, K., Suzuki, M. and
Nakajima, Y.
TITLE CONTROL OF ADVENTITIOUS GENE MANIFESTATION AND PRODUCTION OF
JOURNAL ADVENTITIOUS GENE PRODUCT USING SAID CONTROL
PATENT: JP 1988317088-A 1 26-DEC-1988;
MITSUI TOATSU CHEM INC
COMMENT OS Rhodospiridium toruloides
PN JP 1988317088-A/1
PD 26-DEC-1988
PF 18-JUN-1987 JP 1987152359
PI FUKUHARA NOBUHIRO, YOSHINO SADAO, WATANABE MIDORI, PI
YAMAMOTO KAORU,
SUZUKI MAKI, NAKAJIMA YOSHIYUKI
PC C12N15/00, C12N9/88, (C12N9/88, C12R1.01);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=IFO 559 (ATCC 10788);
FH Key Location/Qualifiers
FT CDS 1..2151
FT 3'UTR 2152..2331.
FEATURES
source Location/Qualifiers
1..2331
/organism="Rhodospiridium toruloides"
/mol_type="genomic DNA"
/db_xref="taxon:5286"

ORIGIN

Query Match 67.8%; Score 1677.4; DB 6; Length 2331;
Best Local Similarity 80.4%; Pred. No. 1.3e-174;
Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;

Qy 1 ATGGCCCCCTCCTGATCGATCGCGACTCGTGTGCGCAACGGGCTGCAAGGATG 60
Db 1 ATGGCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
Qy 61 CAGCGCGNNCCGNCNNNGCNCNACGCGCGCGCAAGTCCACNCTGNCNGCGCGCGNG 120
Db 61 CAGCGCTGT-----CAATGCGCGCTCGACCAACCTC-----GCACTGCGAG 100
Qy 121 GCTCGCTCTCCCGACCAACCGAGNAGCGAGCTCGACATGTGAGNAGATCTCGCGG 180
Db 121 GCTCGCTCTCCCGACCAACCGAGNAGCGAGCTCGACATGTGAGNAGATCTCGCGG 180

Db 101 GCTCGACCTGCCCCCAACCCAGGTCACGAGCTGACATGTGAGAGATCTCGCGG 160
Qy 181 ACCCAACCGGNNACGAGNNNTTGAATCGAGGGGTACACCTTCACCTCGGAGACGTG 240
Db 161 CGCGGACCGACTGCA---CGCTGAACTCGACGGGTACTGTCTCAACTCGGAGAGCTG 217
Qy 241 TGGGGCGCGGCGCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 300
Db 218 TCTCGCGCGGAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
Qy 301 AAGATGCAAGNAGAGTGAAGTCTCGGCTCGGNNCAGCTCNAACAACNGCTACGAGT 360
Db 277 AAGATTGACAAATCGGTGAGATCTGTGGGCTCGGAACTTCACAGACGCTACAGG 336
Qy 361 AGGACTGTTTGGCGGCTCGGCGGACACCGGACTAGAGATGATCTCGTTCAGAG 420
Db 337 AGGACTGATTTGGGGAATCGGACAGACACCGGACAGAGACGCGATCTCGCTCAG 396
Qy 421 GCNCTCTCGAGACACAGACTCTGGCGGTGTCGNNCGTTCGNNCGTTCGNNCGTTC 480
Db 397 GCTCTCTCGAGACACAGACTCTGGCGGTGTCGNNCGTTCGNNCGTTCGNNCGTTC 456
Qy 481 GGNCGCGGCTTCGAGAACTCGCTTCGCTGAGGTCGTCGCGCGCGCATGACATTCG 540
Db 457 GCGCGCGGCTTCGAGAACTCGCTTCGCTGAGGTCGTCGCGCGCGCATGACATTCG 516
Qy 541 GTCACTCGCTACACGCGGCGGCACTCGGNNCTCGGCTCGGCTTCGNNCGGCTCAC 600
Db 517 GTCAACAGCTTGACCGCGGCGGCACTCGGNNCTCGGCTCGGCTTCGNNCGGCTCAC 576
Qy 601 AACTTCTCAACCAACGCGATCACTCGTCCCTCGCGGCAACATCTCGGCGTGC 660
Db 577 AACTTCTCAACCAACGCGATCACTCGTCCCTCGCGGCAACATCTCGGCGTGC 636
Qy 661 GCGGACCTCTCCGCTCTCTCTATATGATGCGCGCGGCTACCGGTACCGGAGNN 720
Db 637 GCGGACCTCTCTCTCTCTCTATATGATGCGCGCGGCTACCGGTACCGGAGNN 696
Qy 721 GTNCACTNNNTNCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
Db 697 GTGACGCTGCTGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 756
Qy 781 TTNGGCTCGAGCCCGCTGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db 757 TTCAACCTCGAGCCCGCTGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816
Qy 841 GCGGCTCGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 817 GCGGCTCGGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876
Qy 901 TCGAGAGGCTCACAGGCTTNAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
Db 877 TCGAGAGGCTCACAGGCTTNAAGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
Qy 961 CANTTCTCCACAGGCTGAGGCGGCTCACCCCAACCAAGTCAAGGTGCGGCAAGT 1020
Db 937 CANTTCTCCACAGGCTGAGGCGGCTCACCCCAACCAAGTCAAGGTGCGGCAAGT 996
Qy 1021 CGCAGGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 997 CGCAGGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
Qy 1081 GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
Db 1057 GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
Qy 1141 GAGCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 1117 GAGCGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
Qy 1201 TCGAGGACGAAACCGGCTCATGACGTCGAGAGCAAGAGACCAACGAGGAGCAAC 1260
Db 1177 TCGAGGACGAAACCGGCTCATGACGTCGAGAGCAAGAGCTTCGAGGAGGAGCAAT 1236

QY 1261 TTCAGAGCGNCGCTGTGCGNAAACAGATGAGAAAGACTCGCCTCGCNCCTTGCCCTGATC 1320
DB 1237 TTCAGAGCTGCGCTGTGGCCAAACATGAGAAAGACTCGCCTCGGCGTGGCCCAATC 1296
QY 1321 GGCAAGCTCAATTGACGAGCTCACGAGATGCTCAACGCGCGCATGAAACCGGCGCTN 1380
DB 1297 GGCAAGCTCAATTGACGAGCTCACGAGATGCTCAACGCGCGCATGAAACCGGCGCTN 1356
QY 1381 CCNTCTGCTGCTGCTGCGAGAGACCCCTGCTCTCTATCACTGAAAGGCGCTGACATT 1440
DB 1357 CCCTCTGCTGCTGCGAGAGACCCCTGCTCTCTATCACTGAAAGGCGCTGACATC 1416
QY 1441 GCGCGNCGNCTAATCACTTGGAGCTCGGACCTTGGCAACCCGCTNACGACCCAGTCT 1500
DB 1417 GCGCGNCGNCTAATCACTTGGAGCTCGGACCTTGGCAACCCGCTNACGACCCAGTCT 1476
QY 1501 CAGCGCGNCGAGATGAGGCAACAGGCGCTCAACTGCTGCGNCTCATCTGCGCGCGCG 1560
DB 1477 CAGCGCGCTGAGATGAGGCAACAGGCGCTCAACTGCTGCGCGCTCATCTGCGCGCTG 1536
QY 1561 ACNCGCGAGGCAAGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
DB 1537 CAGCGCGAGTCAAGAGCT 1596
QY 1621 CAGCGCGTGAAGCTGCGGCGGAGTGGAGTTGAGTTCAAGAAAGAGTTGAGCCGCTNNTC 1680
DB 1597 CAGCGCGTGAAGCTGCGGCGGAGTGGAGTTGAGTTCAAGAAAGAGTTGAGCCGCTATC 1656
QY 1681 NCNCGCTCTNTGAGAGCACTTTGGCNCNCGCCTCGACGCGCNCNNNNNNNNNNNACGAA 1740
DB 1657 GCTGCGCTATGACAGACACTTTGGCTCCGCGCATGACCGGCTGAACTGCGGAGAG 1716
QY 1741 CTCGAGGAAAGGTCAACAAGNCGCTCAAGAGCNCCTGAGAGAGCACTTCTGTAAGAC 1800
DB 1717 CTCGAGGAAAGGTCAACAAGNCGCTCAAGAGCNCCTGAGAGAGCACTTCTGTAAGAC 1776
QY 1801 CTCGAGCGCGCTGAGAGCAAGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1860
DB 1777 CTCGAGCGCGCTGAGAGCAAGCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1836
QY 1861 NNCGTCTGCGCNCNNNNGCCANNAGTCTGCTGCGCNCGCGCTCAAGCGCTGAGAGTCTG 1920
DB 1837 TCGTCAAGCTC-----GCTCTGCGTGGCGCGCTCAAGCGCTGAGAGTCTGCGC 1885
QY 1921 CCGCGGAGAGGCAATCTGCTCAAGCGCNCNNNNCTGCGAGCNCCTCTGAGAGCGCNCCT 1980
DB 1886 CCGCGGAGGCAATCTGCTCAAGCGCNCNNNNCTGCGAGCNCCTCTGAGAGCGCNCCT 1945
QY 1981 CGTCTGCTGCGCGCGCTCNCCTGTAAGTCTCTGCGCGGAGCGCGCTCTCTGTAAGTCTG 2040
DB 1946 CAGACTGCTGCGCGCGCTCTCTGTAAGTCTCTGCGCGGAGCGCGCTCTCTGTAAGTCTG 2005
QY 2041 TCGCGAGAGGCTCGGCGTCAAGGCGCGCGCGCGAGCTCTTCTCTGCGGAGAGAG 2100
DB 2006 TCGCGAGAGGCTCGGCGTCAAGGCGCGCGCGCGAGAGCTCTTCTCTGCGGAGAGAG 2065
QY 2101 TGAAGATCGGCAACAGCTCTCCCGCATATAGAGGCGCTCAAGNCGCGGAGCTCAAC 2160
DB 2066 TGAAGATCGGCTGAGAGCTCTCCAGATCTTGAAGGCGCTCAAGGCGGAGAGTCAACA 2125
QY 2161 ACCTCTCTCTCAAGATGCTCGCTAGAGNCGCNCNNNNCTCTGCTNNNNNNCCNNNNC 2220
DB 2126 ACCTCTCTCTCAAGATGCTCGCTAGAGCTCTTCCCACTCTGCGATCCCTTCATACCC 2185
QY 2221 NNCCNNNNNNNNCTTNNNNNNCTGNNNNNNCTGNNNNNNNNNNNNNNNNNNNNNNNN 2280
DB 2186 TATTCGCGCTGACCTTTAGAGCTGCTTCTGTTGAGACTGGAATCTGCACTGCTTCTT 2245
QY 2281 TNNNNNNNNNNCTNN 2340
DB 2246 TCGTCTGCTGCT 2304

QY 2341 NGTNNNNNNNN 2352
DB 2305 AGTACCATCA 2316
RESULT 8
RTPALGNA
LOCUS RTPALGNA 2315 bp mRNA linear PLN 23-MAR-1993
DEFINITION R.toruloides mRNA for L-phenylalanine ammonia-lyase (PAL).
ACCESSION X12702
VERSION X12702.1 GI:288350
KEYWORDS L-phenylalanine ammonia-lyase; PAL gene.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
REFERENCE 1 (bases 1 to 2315)
AUTHORS Unknown.
JOURNAL Unpublished
COMMENT (Patented) EP 0 260 919 A1.
FEATURES
source location/Qualifiers
1..2315
/organism="Rhodospiridium toruloides"
/mol_type="mRNA"
/strain="IPO 559"
/db_xref="taxon:5286"
/clone="pSW13"
1..2315
/gene="PAL"
1..2151
/gene="PAL"
/codon_start=1
/product="L-phenylalanine ammonia-lyase"
/protein_id="CA31209.1"
/db_xref="GI:288351"
/db_xref="GOA:P11544"
/db_xref="SWISS-PROT:P11544"
/translation="MABSLDISHPANGVASAKQAVNGASTNLVAGSHLPPTQVTC
VDIVKMLAPDSTLELDGYSLNGDYVAARKGKRVYVDSDEISKIDKSEFLR
SOLSWYGVATGTTGGGADRTEDAIKLOKXLEHOLGCVLPSDFSGRLGNEL
PLEYVRGAMTIRVNSLRGSAVAVLVLEALTNLNGITPIVLRGTIASGLISPL
STLAALISGHPDSKYHVHVESEKILYAREMALFNEIPVYLGKSELGLVNGRVA
SMATLALDHAMLSLDSLTAKTVEAMVGHGSHFFLHDVTRBHFQLEVGAKNIK
LLEGRFVHHEEVKVDDEGLRQRYPLRTSPQWGLVSDLIHAHVLTLBAQ
STDPNPLDVENKTSHHGNGFOAAVANTMEKTLGLAIGKUNLFTQLEMLNAGNR
GLPSCLAEDESLSYHCKGDLIAAAVTSSEGLHANVTTTHVCPAEVANAQVNSLAI
SARPTTESNDVLSILATHTVCVLAIDIRLIEEPKRGKGPVAVSIIDHFGSAMTG
SNLRPELYEKYKTLAKRLBOTNSYDLVPRHDAFSRAAGVAVVTSLSLSLAANA
WVAABSAISLTQVRETFMSASTSPALSYLSPRTQILYAVREBELYKARQDV
FLGKGEVLTIGSVSKTYEALIKGRINNVLMKLA"
2315
/gene="PAL"
polyA_site
ORIGIN
Query Match 67.7%; Score 1676.4; DB 8; Length 2315;
Best Local Similarity 80.5%; Pred. No. 1.7e-174;
Matches 1890; Conservative 0; Mismatches 422; Indels 36; Gaps 6;
QY 1 ATGGCCCCCTCNCNTGAGCTGATCGGACCTCGANTGCGCAAGCGCNCNAAAGGNTG 60
DB 1 ATGGCACTCTCCTGATCGATCTTCGCACTCGTTTCGAAAGCGCGTGCATTCGCGAAAG 60
QY 61 CAGCGCNCNCNCNNNNGCNCNAAGCGGCGCGCACGCTCACTGNCCTGNGCGCGCNCNG 120
DB 61 CAGGCTGT-----CAATGCGCGCTCGACCAACTC-----GAGTTCGAG 100
QY 121 GCTGCTCTCTCCGACCAACCAAGNNGAGAGCTGACATGTTGAGAGATCTCTCGCG 180
DB 101 GCTGCACTGCGCCCAACCAAGCTGACGAGGTGACATGCTGAGAAAGATGCTCGCG 160
QY 181 ACCCAACCCGNNACGAGCNCNTGAACTGACGAGTACACCTCAACCTCGAGNACGCTG 240
DB 161 CGCGACGCACTGCA---CGCTGAACTGACAGGCTACTCTGCTCAACTCGGAGACGCTG 217

QY 241 TCAGCGCGCGNCGCAAGAGGCGCGNCGTCCGCGTGCAGACAGNCGACGAGATCCGCGCA 300
Db 218 TCTCGCGCGGAGAGAGGCGAGGCGCTGTCCGCGTCAAGACAG-CGACGAGATCCGCTCA 276
QY 301 AAGATCGACAAAGAGNGTCGAGTTCTTCGNNCNCACTCMACAAAGNGTCACGNGTC 360
Db 277 AAGATTGACAAATCGGTGCGAGTTCTTGCGCTCGCAACTCTCATAGAGGTCA CGGCGTC 336
QY 361 ACCGACTGGTTGCGCGGCTCGGCGGACACCGCGGAGTGGAGTGCMACTCCGCTCCAGAG 420
Db 337 ACGACTGAGATTGCGGAGATCGGACACCGCGACGAGAGGCCACTCGGCTCCAGAG 396
QY 421 GCNCTCTCGAGCAGCAGCTCTGCGGTGTCGCCNACGTCGNTGANTCCCTTCGCGCTC 480
Db 397 GCTCTCTCGAGCAGCAGCTCTGCGGTGTCGCCNACGTCGNTGANTCCCTTCGCGCTC 456
QY 481 GAGCGCGGCTCGAGACTCGCTTCGCTCGAGGTGTCGCGGCGGCGCATGACATCCGC 540
Db 457 GCGCGGCGTTCGAGAACTCGCTTCGCCCGAGGTGTCGCGGCGGCGCATGACATCCGC 516
QY 541 GTCAACTCGCTCAACGCGGCGCACTGCGNGTCCGCGCTGTCGTCCTCGAGGCGCTCAC 600
Db 517 GTCAAGAGCTTGACCGCGGCGCACTGCGGTGTCGCGCTGTCGTCCTCGAGGCGCTCAC 576
QY 601 AACTTCTCAACACGAGCATCAACCCCATGTCCTCCCTTCGCGGCGCATGTCGAGCTCG 660
Db 577 AACTTCTCAACACGAGCATCAACCCCATGTCCTCCCTTCGCGGCGCATGTCGAGCTCG 636
QY 661 GCGACCTCTCCCGCT 720
Db 637 GCGACCT 696
QY 721 GTNACGTNNTCACGAGGCGAGNAGAGATCATGTCGCCCGGAGCGCATGCGCTC 780
Db 697 GTSCAGTGTGTTCACGAGGCGAGAGAGATCTGTAGCGCCGCGAGCGCATGCGCTC 756
QY 781 TTMGTCTGAGCGCGCT 840
Db 757 TTCAACTCGAGCGCGCT 816
QY 841 GCGCTCTCGCGCTCGATGCGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 817 GCGCTCTCGACATCGATGCGACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 876
QY 901 TCGAGAGCGCTCAAGCGCTCTMAAGGTGAGGCGCATGTGCGGCGACCGCGCTCTCT 960
Db 877 TCGAGAGCGCTCAAGCGCGCTCTMAAGGTGAGGCGCATGTGCGGCGACCGCGCTCT 936
QY 961 CCNTTCTCAACGAGTCAAGCGCGCTCAACCGGATCGAGTGGCGCGCAATC 1020
Db 937 CCCTTCTTCAAGAGTCAAGCGCGCTCAACCGGATCGAGTGGCGCGCAATC 996
QY 1021 CGGACGCTCTCGAGGCGAGCANNGTTTGGCGTCAACGAGAGAGGTCAAGTCAAG 1080
Db 997 CGGACGCTCTCGAGGCGAGCANNGTTTGGCGTCAACGAGAGAGGTCAAGTCAAG 1056
QY 1081 GAGGAGAGAGGAGATTTCTCGGCGAGAGCGGCTTACCGGCTCGGCGAGTGGCGTCT 1140
Db 1057 GAGGAGAGAGGAGATTTCTCGGCGAGAGCGGCTTACCGGCTCGGCGAGTGGCGTCT 1116
QY 1141 GCGCGCGCTCGTCAAGCATGATTCAGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCT 1200
Db 1117 GCGCGCGCTCGTCAAGCATGATTCAGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCT 1176
QY 1201 TCGACGACGAGCAACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
Db 1177 TCGACGACGAGCAACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1236
QY 1261 TTCAGAGCGNCGGCTGTGCGNACACGATGAGAGAGACTCGCTCGNCTCGCGCTGATC 1320
Db 1237 TTCAGAGCTGCGCGCTGTGCGCAACACATGAGAGAGACTCGCTCGCGCTCGCGCATC 1296

QY 1321 GCGAAGCTCAACTTCACGAGCTACCGAGATGCTCAACGCGCGGATGAACCGCGGCTN 1380
Db 1297 GCGAAGCTCAACTTCACGAGCTACCGAGATGCTCAACGCGCGGATGAACCGCGGCTN 1356
QY 1381 CCNTCTGCTCTGCTGCGGAGAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
Db 1357 CCCTCTGCTCTGCGGCGGAGAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1416
QY 1441 GCGCGNCGNCTPACTTCTGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTG 1500
Db 1417 GCGCGCTGCGGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTGAGGAGCTG 1476
QY 1501 CAGCGCGNAGATGAGGAGCAACGAGCGGTCATCTGCGNCTCATCTGCGNCGCGC 1560
Db 1477 CAGCGCGCTGAGATGAGGAGCAACGAGCGGTCATCTGCGNCTCATCTGCGNCGCGC 1536
QY 1561 ACGCGCGAGGCGCAAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Db 1537 ACGACTGAGTCCAAAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
QY 1621 CAGCGCGTCACTCTCGCGGAGTTCAGGTTCAAGAGAGATTGCAACCGGANTTTC 1680
Db 1597 CAGGCGATCGACTTGGCGGAGTTCAGGTTCAAGAGAGATTGCAACCGGANTTTC 1656
QY 1681 NCGNCGCT 1740
Db 1657 GTCTGCTCATCGACGAGCACTTGGCTCGCGCTGACCGGCTGACCTGCGCGAG 1716
QY 1741 CTCGAGCAAGAGTCAACGAGNCTGAGAGAGCAACCTGTCGAG 1800
Db 1717 CTCGAGAGAGTCAACGAGNCTGAGAGAGCAACCTGTCGAG 1776
QY 1801 CTCGAGCGGCTGAGCAAGCTTCTGTTGCGACCGGCAACCTGTCGAG 1860
Db 1777 CTCGAGCGGCTGAGCAAGCTTCTGTTGCGACCGGCAACCTGTCGAG 1836
QY 1861 NNGTCTGCGGNNNCGANNAGTCTGCTCGGCGGCTGCAACGCTGGAAGTGGCT 1920
Db 1837 TCGTCAAGTCT-----GCTCTGCTGCGGCGGCTGCAACGCTGGAAGTGGCT 1885
QY 1921 CCGCGAAGAGGAGCTCTGCTCAAGCGGAGGTCGCGCAACCTTCTGAGCGGCGCT 1980
Db 1886 CCGCGAAGTGGCGATCTGCTCAACCGGAGGTCGCGCAACCTTCTGAGCGGCGCT 1945
QY 1981 GGTGCTGTCGCGCGCTCNGTACTCTTCTGCGGCGAGCGGCTGCTGACTCGTTG 2040
Db 1946 GAGCTGCTGCGCGCGCTCTGCTGCTGCGGCGAGCTGAGATCTCTTACGCTTCTG 2005
QY 2041 TCGCGAAGAGCTCGGCGTCAAGGCGCGCGGCGAGCTTCTCTGCGCAAGCAGAG 2100
Db 2006 TCGCGAAGAGCTTGGGCTCAAGGCGCGCGGCGAGCTTCTCTGCGCAAGCAGAG 2065
QY 2101 TGAAGATGAGCAACAGTCTCCGATCTAGAGGCGCATCAAGGCGAGGCGGCGCATCA 2160
Db 2066 TGAAGATGAGTCAAGGCTCTCAAGATCTAGAGGCGCATCAAGGCGAGGCGGCGCAT 2125
QY 2161 ACGTCTCTGTAAGATGCTGCTGAGACCTTCTTCTCACTCTGCACTCCCTTCATACC 2220
Db 2126 ACGTCTCTGTAAGATGCTGCTGAGACCTTCTTCTCACTCTGCACTCCCTTCATACC 2185
QY 2221 NNNCGNNNNNNNTTNGNNNTGNNNTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2280
Db 2186 TATCCGCGCTGCACTTGAAGACTGCTTCTTGGAGACTGAGTCTCGAGTCTGCTTCT 2245
QY 2281 TNNNNNTNNNTNNCTCNCNNNNNANNGTNNNTNNNNNTNNNNNTNNNNNNNNNN 2340
Db 2246 TCGTTCTTGGTGGCTCTC-TAGACGCTGCGGTATTACTCGAGATTGTGAATACAGC 2304
QY 2341 NGTNNNCA 2348
Db 2305 AGTACCA 2312

Db 1816 CTGAGGCGGCTGGGACGACGCTTCTGTACGACCGGACCGCTGTGAGCTCTC 1875
 1861 NNNGCTCGCCNNNNGCANNAGTCTGCTCGCCGCGCTCAACGCTGGAAGTGGCT 1920
 1876 TCGTCTCGCCCTCTGCA--ACGTACCCCTTACTGCGCTCAACCGCGGAGGTTGGCT 1933
 1921 CCGCCGAGAAAGCCATCTGCTCAACGCGCANGTCGCGACNCCTTTCGACGCGCCGT 1980
 1934 CGGCGGAGAAAGCCATCTGCTCAACGCGGAGAGTGGCAACGCTTTCGAGAGACGCGGT 1993
 1981 CGGCTGTCGCGCCGCGCTCNCGTACTCTCGCCGCGGACGCGCGCTCTGTACTGTTGG 2040
 1994 CTTCGAGGCGCGCGGACGACGATCTCTCGCCGCGGACGCGCGCTCTGTACTGTTGG 2053
 2041 TCCGCGAGAGCTCGGCGCTCAAGGCGCGCGGCGGACGCTCTCTCTCGCAAGAGAGG 2100
 2054 TGGCGGAGAGCTCGGCGGTGAGGCGCGCGCGGAGAGTGTGTTGTCGCGGTGGAGAGG 2113
 2101 TGAAGATCGGACCAACGCTCTCCGCTACTAGAGGCGCATCAAGNCGCGCATCAACC 2160
 2114 AGAGATCGGAGACCAACGCTCTCGGCTACTAGAGGCGCATCAAGGACGCGCATCAACC 2173
 2161 ACGCTCTGCTCAAGATGCTCGTGTAGNNCNCNNNANNCTCGCNTNNNNCCNNNCC 2220
 2174 ACGTCTCTGCTCAAGATGCTCGGTA--AGGCCGAGAGACCTCGCTAGAGCGCCGCTC 2232
 2221 NNNCCNNNNNNCTTNGNNNTGNNTCNTGNCNNNNCCGANNNNCCNNNNNN 2280
 2233 ACCCGAAGACGACCTTTTGCAGCGTGTGTGCAAGACGAGCTTCTCTCATACACA 2292
 2281 TNNNNCTNNCTNNCTCNCNNNNANNCTNNNNCTNNNNCTNNNNCTNNNNNNNN 2340
 2293 TGTGCGCTTACTCTGCGCGCTGATCAGCTCTCTCACTTCTTTCGTAATCCGCGCTCTC 2352
 2341 NGTNNCCANNACNCNTNNNNNNANNCCGANNANGANNANTNNANNNTGANNNNNN 2400
 2353 GGTGCTCA--GTACACGTGTATAGAGCTGGAATGATGCAAGCTTTCGAGTTCAAAAA 2411
 2401 NNNNAAAA 2408
 2412 AAAAAA 2419

RESULT 10
 AX36860 2419 bp DNA linear PAT 15-FEB-2002
 LOCUS AX36860
 DEFINITION Sequence 12 from Patent WO0208402.
 ACCESSION AX36860
 VERSION AX36860.1 GI:18698149

SOURCE Rhodotorula graminis
 ORGANISM Rhodotorula graminis
 Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
 Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
 Rhodotorula.

REFERENCE 1
 AUTHORS Yoshida, R.K. and Koester, A.B.
 TITLE Phenylalanine ammonia lyase polypeptide and polynucleotide
 JOURNAL sequences and methods of obtaining and using same
 Patent: WO 0208402-A 12 31-Jan-2002;
 PCB Services, Inc. (US)
 Location/Qualifiers

FEATURES
 source 1..2419
 /organism="Rhodotorula graminis"
 /mol_type="unassigned DNA"
 /db_xref="taxon:29898"
 37..2199
 /note="unassigned protein product"
 /codon_start=1
 /protein_id="CAD23828.1"
 /db_xref="GI:18698150"
 /db_xref="REMBTREMBL:CAD23828"

/translation="MADSLSLATTANGTNGSHAAPTSAAAPTSALRRTPGLDGH
 AAHQSLIVELISDPDIDVELSGSLVRLDVGAARGRVAVQNDDELIRADK
 SYDFLQLONSVYVGTTFRGSGADTPELVSLOKALIDHQLPFTPSXSPFVGR
 GLENTLPLEVVRGAMVIRVNSLTGSHAVLVLEALTNLRIRTPVLRGSLIAS
 GDLSPSLYAGAITGHPDVKRVHLEETKEKIMFARBAISLFGELAVLGPREGGLVN
 GRVYASNAITSLIDSHMLSLISQALTLALVEAMVGOGSGFAPRIHVCVRPHGQVEV
 ARNRTTLSSGSFVHEHEEKVYDDEGTRIPDRIPRTSPQSLPELVEMHAYISL
 SLENNITDNLPLEVENKQTAHGNFPAASVVISMEKTRIALILGILNTOCTBLN
 AAMRGSPSCIAEDPSLNYHKGJLDIHAAYASLEGHLLNPVTVPQPAEMGNQVN
 SLALISARITAEADVLSLILASHYLCLOADVLRAMELDFKQFDPILPTLQCHLG
 TGLDYNALALEKVALNKRLPQETTYLEPRMDAPFAYATGVVELSSPSANVTLT
 AVNMKVASAEKALISLIREVNRPMOPSSADAPAHVLSFPRFVLVSFVREELGVGAR
 RGVDFVQDETIGSNVRIYBAIKDRINHVLYKMLA"
 493..495
 /note="Other information: Xaa = Val or Ala"

ORIGIN

Query Match 65.8%; Score 1629; DB 6; Length 2419;
 Best Local Similarity 78.0%; Pred. No. 2,58-169;
 Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;

1 ATGGCCCCCTCTCTGACTCGATCGCACTCGTGCACCAAGCGCNTCNAAAGCNTNG 60
 37 ATGGCCCCCTCTCTGACTCGATCGCACTCGTGCACCAAGCGCNTCNAAAGCNTNG 96
 61 CAGCGCGGCGCCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
 97 CAGCGCGGCTCGGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 154
 121 GCTGCTCTCTCTCGGACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 155 GCTGATAGGCG 214
 181 ACCCGACGNNACGACGNNNTTCAACTCGAGGCTCACCTTCACCTCGAGAGAGTGG 240
 215 ACCCGACGAC--GACGTCTGAGGCTCAGCGGGTCAACCTTCACCTCGAGAGTGG 271
 241 TGGGCGCGCGCGCAAGGCG 300
 272 TGGGCG 330
 301 AAGATCGACAAAGNGTGAAGTCTCTCGGNNCAGCTCAACAACAAGNCTTACGNGTC 360
 331 CGGTCGACAAAGAGGTGAGCTTCTCAAGGCCAGCTTCAGAACTCGGTCACGAGTTC 390
 361 ACGATGATTCG 420
 391 ACCAGGAGTTTGGTGGCTCGGCGGACACGAGACTAGAGATGAGTCAAGCTTCAGAG 450
 421 GCNCTCTCGAGACCAAGCTTGGGTTGTCGCCACGTCGNTCGANTCGTTCGCGCTC 480
 451 GCGTCATCGAGACACAGCTTGGGCGGTGACGCGCGAGTCGCTGCTTCAGGATC 510
 481 GCGCGCGCGCTCGGAACTCGCTTCGCTGAGGTCTCGGCGCGCGCGCGCGCGCGCGCG 540
 511 GAGCGCGGCTCGGAAACAGCTTCGCTGAGGTCGTCGCGCGCGCGCGCGCGCGCGCG 570
 541 GTCACTGCTCAGCG 600
 571 GTCACTGCTCAGCG 630
 601 AACTTCCTCAACAGCGGATCAACCCCATGATGCTCCCTCGCGCGCGCGCGCGCGCGCG 660
 631 AACTTCCTCAACAGCGGATCAACCCCATGATGCTCCCTCGCGCGCGCGCGCGCGCGCG 690
 661 GCGGACCTCTCG 720
 691 GCGGACCTCAAGCGCGCTCTGTAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
 721 GTNACGTTNTTCAAGAGGCGCANNAGAGATCATGTGCGCGCGCGCGCGCGCGCGCGCG 780
 751 GTTCACTGTTTGAACGAGGAGAACGAGAGATCATGTTTGGCGCGCGCGCGCGCGCGCG 810

QY 781 TTNGCTCTGAGCCCGCTGCTCTCGGCCCCGAGAGAGGCTCTCGGCTCTGTCACGCGCAG 840
DB 811 TTGTGCTCGAGGCGAGTGTCTCTCGGCCCGAAGAGGGTCTCGGCTGTGTCACGAGCAG 870
QY 841 GCCGCTCTCCGCTCGATGCGCAGCCCTCGCTCTGCACGACGCACATGCTCTGCTCTC 900
DB 871 GCCGCTCTCCGCTCGATGCGCAGCCCTCGAGTGTGCAAGTGTGCAATGCTCTGCTCTC 930
QY 901 TCGCAGGCGCTCAGCGGCTCTGACGAGGCAATGCTGCGGCGCAGCGCGCTGCTTCCAC 960
DB 931 TCGCAGGCGCTCAGCGGCTCTGACGAGGCAATGCTGCGGCGCAGCGGCTGCTTCCAC 990
QY 961 CCNTTCTCTGACGAGCTGACGCGCTTCAACCGCAGCCGATGAGGTGCGCGCAATC 1020
DB 991 CCGTTCATCAGAGGCTGCTGCGCGCGCGCAGCCCGCGCAGGTGAGGTGCGCGCAATC 1050
QY 1021 CGCAGGCTCTGAGGCGAGCGGCTGCTGCGCAGCCGATGAGGTGAGGTGAGGTGAGGT 1080
DB 1051 CGCAGGCTCTGAGGCGAGCGGCTGCTGCGCAGCCGATGAGGTGAGGTGAGGTGAGGT 1110
QY 1081 GACGACGAGGCGATTCCTCGCGCAGGACCGCTACCGGCTCGCAGGCTGCGCTCAGTGCTC 1140
DB 1111 GACGACGAGGCGATTCCTCGCGCAGGACCGCTACCGGCTCGCAGGCTGCGCTCAGTGCTC 1170
QY 1141 GCGCGCTCTGTCAGGCGACATGATTTACGCGCCACCGGCTCTGCTGAGGCGCGCAG 1200
DB 1171 GCGCGCTCTGTCAGGCGACATGATTTACGCGCCACCGGCTCTGCTGAGGCGCGCAG 1227
QY 1201 TCGACGACCGCAGACCGCGCTCATGACGTCGAGAACGAGAGACCGAGCGCGCAGC 1260
DB 1228 ACGACGACCGCAGACCGCGCTCATGACGTCGAGAACGAGAGACCGAGCGCGCAGC 1287
QY 1261 TTCGACGCGCGCTGTCGCGCAGAACGATGAGAAAGTCTGCTGCGCTGCGCTGATC 1320
DB 1288 TTCGACGCGCGCTGTCGCGCAGAACGATGAGAAAGTCTGCTGCGCTGCGCTGATC 1347
QY 1321 GCGACGCTCACTTACGACGCTCACGAGAGTCTCACGCGCGCGCAGAACCGCGCGCT 1380
DB 1348 GCGACGCTCACTTACGACGCTCACGAGAGTCTCACGCGCGCGCAGAACCGCGCGCT 1407
QY 1381 CCNTCTGCTGCTGCTGCGAGAGACCGCTGCTCTGCTCTGCTGCTGCTGCTGCTGCT 1440
DB 1408 CCGTCTGCTGCTGCTGCGAGAGACCGCTGCTCTGCTCTGCTGCTGCTGCTGCTGCT 1467
QY 1441 GCGCGCGCGCTGCTGCTGCGAGAGACCGCTGCTCTGCTCTGCTGCTGCTGCTGCT 1500
DB 1468 CACATCTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTTACGCTT 1527
QY 1501 CAGCGCGCGAGATGAGGCGACACGAGCGCTCAACTGCTGCGCTCATCTGCGCGCGC 1560
DB 1528 CAGCGCGCGAGATGAGGCGACACGAGCGCTCAACTGCTGCGCTCATCTGCGCGCGC 1587
QY 1561 ACNCGCGAGCGCAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
DB 1588 ACTGCGCGAGCGCAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1647
QY 1621 CAGCGCGCGAGCTGCGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1680
DB 1648 CAGCGCGCGAGCTGCGCGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1707
QY 1681 NCNCGCTCTNAGCAGACTTGTGCGCGCGCTGAGCGCGCGCGCGCGCGCGCGCG 1740
DB 1708 CCGAGCT 1755
QY 1741 CTGCGNGAGAAGGTCAACAAGCGCTCAACAAGCGCTCTGAGCAGACGACGCTTACGAC 1800
DB 1756 CTGCGCGCTGAGGTCAAGAAGCGCTCAACAAGCGCTCTGAGCAGACGACGCTTACGAC 1815
QY 1801 CTGAGCGCGCGCTGAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1860
DB 1816 CTGAGCGCGCGCTGAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1875
QY 1861 NNCTCTCTGCT 1920

DB 1876 TGTGCTGCGCGCTCTGCGCA--ACGTCAACCTTACTGCGCTCAACGCGTGAAGGTGCT 1933
QY 1921 CCGCGGAGAAGGCGATCTGCTGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
DB 1934 CCGCGGAGAAGGCGATCTGCTGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1993
QY 1981 CCGTCTGCTGCG 2040
DB 1994 CTTGCGAGCG 2053
QY 2041 TCGCGGAGAAGTCTGCGCTGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
DB 2054 TCGCGGAGAAGTCTGCGCTGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2113
QY 2101 TCGAGATGCGCAGCAAGCTCTCCGCGCTTACAGAGCGCATCAAGGCGCGCGCGCG 2160
DB 2114 AGAGATGCGGAGCAGCTCTGCGCGCATCTAGAGCGCGCATCAAGGCGCGCGCGCG 2173
QY 2161 ACCTCTGCTGCAAGTCTGCGCGTGTGCGCGTGTGCGCGCGCGCGCGCGCGCGCG 2220
DB 2174 ACCTCTGCTGCAAGTCTGCGCGTGTGCGCGTGTGCGCGCGCGCGCGCGCGCGCG 2232
QY 2221 NNCG 2280
DB 2233 ACCCGAGACAGCTTTCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2292
QY 2281 TNNCG 2340
DB 2293 TGTGCG 2352
QY 2341 NCTNNCG 2400
DB 2353 GGTGCTCA-GTACAGCTGTATGAGCTGTGAGATGAGATGAGATGAGATGAGAT 2411
QY 2401 NNNANAAA 2408
DB 2412 AAAAAAAA 2419

RESULT 11
E01759
LOCUS E01759 2151 bp RNA linear PAT 29-SEP-1997
DEFINITION DNA encoding L-phenylalanine ammonia-lyase.
ACCESSION E01759
VERSION E01759.1 GI:2170012
KEYWORDS JP 198291583-A/1.
SOURCE Rhodospiridium toruloides
ORGANISM Rhodospiridium toruloides
REFERENCE 1 (bases 1 to 2151)
AUTHORS Fukuhara,N., Yoshino,S., Yamamoto,K., Nakajima,Y., Iwata,T.,
Watanabe,M., Suzuki,M. and Makiguchi,N.
TITLE AMINO ACID SEQUENCE OF L-PHENYLALANINE AMMONIA LYASE, ITS
STRUCTURAL GENE, NOVEL BASE SEQUENCE CONTAINING SAME, TRANSFORMANT
PRODUCED THEREWITH AND PRODUCTION OF L-PHENYLALANINE USING SAME
JOURNAL Patent: JP 198291583-A 1 29-NOV-1988;
MITSUBI TOASTU CHEM INC
COMMENT OS Rhodospiridium toruloides
PN JP 198291583-A/1
PD 29-NOV-1988
PF 22-MAY-1987 JP 1987123950
PI FUKUHARA NOBUHIRO, YOSHINO SADA0, YAMAMOTO KAORU, PI
NAKAJIMA YOSHIYUKI,
PI IWATA TOMOYUKI, WATANABE MIDORI, SUZUKI MAKI, PI MAKIGUCHI
NOBOYOSHI
PC C12N9/88,C12N1/16,C12N1/18,C12N1/20,C12N15/00,C12P13/22, PC
(C12P13/22,
PC C12R1/19),(C12P13/22,C12R1/865);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;

CC anti-sense: No;
 CC *source: strain=IFO 559;
 CC *source: clone=DSM101;
 FH Key Location/Qualifiers
 FT CDS 1..2151
 FT /product='L-phenylalanine ammonia-lyase', FT
 mat_peptide 1..2148
 Location/Qualifiers
 1..2151
 /organism="Rhodospirillum rubrum"
 /mol_type="genomic RNA"
 /db_xref="taxon:5286"

ORIGIN

Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 56-169;
 Matches 182; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGCCCCCTCCNTGCACTGATTCGGACCTTCGTCGCAACGGCTCNCNAAGGNTNG 60
 Db 1 ATGGACACCTCGCTCGACTCGATCTCGCACTGTCGCAACGGGCTGCAATCCGCAAG 60
 QY 61 CAGCGCGNCCGNNNGNCGNNAAGGGCGGCAAGTCCACNCTCNGCGGCGGCGG 120
 Db 61 CAGGCTGT-----CATGGGCTTCGACCAACCTC---CGAGTCGAG 100
 QY 121 GCTGCTCTCTCCGACACCCAGANNAGCGCAGCTCGATCGTNGAGNAGATCTTCGCG 180
 Db 101 GCTGACCTCGCCACCAACCCAGATCAGCAGCTCGATCGTNGAGNAGATCTTCGCG 160
 QY 181 ACCCCAGCGNAGNAGNAGNAGTGAATCGAGGCTACACCTTCACCTCGGNAAGTTCG 240
 Db 161 CGCGACCACTGA---CGTGAATCGAGGCTACTCGTCAACTCGTGAACCTCGAAGCTG 217
 QY 241 TCGGCGCGCCGCGCAGGCGCGCCGCGCTCGCTCNCAGACAGNAGAGATCCGCGCA 300
 Db 218 TCGGCGCGCGGAGGAGGCGCGCTCGCTCGCTCNCAGAGAG---GAGAGATCGCTCA 276
 QY 301 AAGATCGACAAAGAGTGAATTCCTCCGANNCCAGCTCMAACAAGGCTTACGAGTTC 360
 Db 277 AAGATTGACAAATCGGTGAGATTCCTGCGCTCGCAACTCGTGAAGGCTTACGAGGCTC 336
 QY 361 ACGACTGTTTGGCGGCTCGGCGCAGACCCGCACTGAGATGAGNAGATCCGCGCA 420
 Db 337 ACGACTGATTTGGCGGATTCGCGACACCCGCACTGAGATGAGNAGATCCGCGCA 396
 QY 421 GCGCTCTCGAGACCAAGCTTCGCGGTGTCCTCCNACGTCGANTCGTCTTCGCGCTC 480
 Db 397 GCTCTCTCGAGACCAAGCTTCGCGGTGTCCTCCNACGTCGANTCGTCTTCGCGCTC 456
 QY 481 GAGCGGCGCTCGAGAACTCGCTCGCTCGAGGCTCGCGGCGCGCATCAATCCGCG 540
 Db 457 GAGCGGCGCTCGAGAACTCGCTCGCTCGAGGCTCGCGGCGCGCATCAATCCGCG 516
 QY 541 GTCACTCGCTCAACGCGGCGCACTCGGAGTTCGCTCGATCTCGAGCGCTCAAC 600
 Db 517 GTCACTCGCTCAACGCGGCGCACTCGGAGTTCGCTCGATCTCGAGCGCTCAAC 576
 QY 601 AACTTCTCAACCAAGGCACTCAACCCCATTCGCTCCGCGGCGCACTTCGCGGCTCG 660
 Db 577 AACTTCTCAACCAAGGCACTCAACCCCATTCGCTCCGCGGCGCACTTCGCGGCTCG 636
 QY 661 GCGGACTCTCCGCTCTCTCAATTCGCGCGGCGCATCAAGGTCACCCGCAACNNCAAG 720
 Db 637 GCGGACTCTCTCTCTCTCTCAATTCGCGCGGCGCATCAAGGTCACCCGCAAC 696
 QY 721 GTTACAGTNNTCAGAGGCGGCAAGAGATCATGTCGCGCGGAGGCGATTCGCGCTC 780
 Db 697 GTTACAGTNNTCAGAGGCGGCAAGAGATCATGTCGCGCGGAGGCGATTCGCGCTC 756
 QY 781 TTNGTCTCGAGCGCGTCTCTCGCGCCGAAAGAGGAGTCTGCTCTGTCACAGGAGC 840

Db 757 TTCAACTCGAGCCGCTGCTCTCGGCGGCAAGAGAGTCTCGATCTGTCACAGGAC 816
 QY 841 GCGGCTCGGCTCGATGAGGAGACCTCGCTCTGAGAGGAGACATGCTCTGCTCTC 900
 Db 817 GCGGCTCGGCTCGATGAGGAGACCTCGCTCTGAGAGGAGACATGCTCTGCTCTC 876
 QY 901 TCGAGGCGCTCAAGGCTCTNACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
 Db 877 TCGAGGCGCTCAAGGCTCTNACGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
 QY 961 CCGTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
 Db 937 CCGTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 996
 QY 1021 CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
 Db 997 CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
 QY 1081 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
 Db 1057 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
 QY 1141 GCGCGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
 Db 1117 GCGCGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
 QY 1201 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
 Db 1177 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1236
 QY 1261 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
 Db 1237 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1296
 QY 1321 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
 Db 1297 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
 QY 1381 CCGTCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
 Db 1357 CCGTCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1416
 QY 1441 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
 Db 1417 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
 QY 1501 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
 Db 1477 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536
 QY 1561 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
 Db 1537 ACNCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1596
 QY 1621 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
 Db 1597 CAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1656
 QY 1681 NCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
 Db 1657 GCTGCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1716
 QY 1741 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
 Db 1717 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1776
 QY 1801 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 Db 1777 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1836
 QY 1861 NCGTCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920

Db 1837 TCGTCGACGTC-----GCTCTGCTGCGCCGCGTCAAGSCCTGGAAGTTCGCG 1885
QY 1921 CCGCCGAGGAAGGCATCTCTGCTCAGCGCGGANGTTCGGACNCTTTGAGCGGACNCTG 1980
Db 1886 CCGCCGACAGTCGACATCTGCTCACCCTGGCAAGTCCGGAGACCTTCTGGTCCGCGGT 1945
QY 1981 CGTCGCTGCGCCGCGCTCNCGTACTCTGCTGCGCGGACGCGCTCTGACTGCTTCG 2040
Db 1946 CGACCTCTGCGCCGCGCTCTGTACTCTGCGCGGACCTCAAGATCTTACCGCTTCG 2005
QY 2041 TCGCGAGAGAGCTCGCGCTCAAGGCTCCGCGGACGCTCTTCTCGGCAAGAGAGG 2100
Db 2006 TCGCGAGAGAGCTGCGCTCAAGGCTCCGCGGACGCTCTTCTCGGCAAGAGAGG 2085
QY 2101 TGAAGATCGGACCAAGCTCTCCGCTCTAGAGGACATCAAGNCGGACGCTCAAC 2160
Db 2066 TGAAGATCGGCTGGAAGCTCTCAAGATCTAGAGGACATCAAGTGGGGCAGATCAAC 2125
QY 2161 AGCTCTCTGTCAGATGCTCGGNTAG 2186
Db 2126 AGCTCTCTCTCAAGATGCTCGCTTAG 2151

RESULT 12
AR205032 2151 bp DNA linear PAT 20-JUN-2002
LOCUS AR205032
DEFINITION Sequence 9 from patent US 6368837.
ACCESSION AR205032
VERSION AR205032.1 GI:21502510
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2151)
AUTHORS Gatenby, A.A., Satriasanti, S., Tang, X.-S., Qi, W.Wei, and Vannelli, T.
TITLE Bioproduction of para-hydroxycinnamic acid
JOURNAL Patent: US 6368837-A 9 09-APR-2002;
FEATURES
source 1..2151
Location/Qualifiers
/mol_type="unassigned DNA"
ORIGIN
Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
Best Local Similarity 83.7%; Pred. No. 5e-169;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;
QY 1 ATGGCCCTCCCTCCTGAGCTGAGTCTGAGACCTGAGTCCGCAACGCGTCCGCAAGCTG 60
Db 1 ATGGCACTCTCGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 60
QY 61 CAGCGCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 120
Db 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GCAATGGCAG 100
QY 121 GCTTCCTCTCTCCGACCAACCAAGNNNGAGCGAGTCTGAGTCTGAGTCTGAGTCTGAG 180
Db 101 GCTTCGACCTGCGACCAACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 160
QY 181 ACCCGACCGGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNN 240
Db 161 CGCGGACCGGACCTGA---CGCTCGAACTGAGCGGCTACTCGTCACTCGTCACTCGTCA 217
QY 241 TCGCGCCCGGCGCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCA 300
Db 218 TCTCGCGCGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
QY 301 AAGATGCAAAANGTGAAGTTCCTCGGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCC 360
Db 277 AAGATGCAAAATGAGTCAAGTCTTGGCTGCAACTCTCTCATAGAGGCTTACGGGCTC 336
QY 361 ACGACTGTTTGGCGGCTGCGCGGACACCGGAGTGAAGTGCNATCTGCTCAAGAG 420
|||||

Db 337 ACGACTGATTTTGGCGGATTCGAGACACCCGCAOCCGAGAGCGGATCTGCTCAGAG 396
QY 421 GGNCTCTCGAGACCAAGCTCTGGGCTGNTCTCCGACGTCGNTGANTCTTCCGCTC 480
Db 397 GCTCTCTCGAGACCAAGCTCTGGGCTGNTCTCCGACGTCGNTGANTCTTCCGCTC 456
QY 481 GGNCGGGGCTCGAAGCTCGGCTCGAGTCCGCTCGAGGTCGCGGCGGCGGATGACCATCCG 540
Db 457 GCGCGGCTCTCGAAGCTCGGCTCGAGTCCGCTCGAGGTCGCGGCGGCGGATGACCATCCG 516
QY 541 GTCAACTGCTCAACGCGGACCTGCGGNTCCGCTCGTCTGTCGAGCGCTCAC 600
Db 517 GTCAAGCTGACCCGCGGACCTGCGGNTCCGCTCGTCTGTCGAGCGCTCAC 576
QY 601 AACTCTCTCAACAGGATGACCCGATGTCGTCGCTCGGCGGACCATGTCGCGCTC 660
Db 577 AACTCTCTCAACAGGATGACCCGATGTCGTCGCTCGGCGGACCATGTCGCGCTC 636
QY 661 GCGGACCTCTCCGCTCTCTCATGCTGCGCGGACATCAAGGTCACCGGACNNCAAG 720
Db 637 GCGGACCT 696
QY 721 GTTCAGGTTTTCAGAGGCGGANNAGAAATCATGTTTTCGCGCGGAGGATGCGCTC 780
Db 697 GTTCAGGTTTTCAGAGGCGGAAAGAAATCTGTAAGCGCGGAGGAGGATGCGCTC 756
QY 781 TTNGCTCTGAGCCCGCTGCTCTCGGCGGAGAGGCTCGGCTCTGTCGTCAGGAGCAG 840
Db 757 TTCAACTCTGAGCCCGTCTGCTCTCGGCGGAGAGGCTCTGCTCTCTCTCTCTCTCTCT 816
QY 841 GCGCTCTCGCTGATGAGGACCTCTCTGTCAGACGACACATGCTCTGCTCTCTC 900
Db 817 GCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 876
QY 901 TCGAGGCGGCTGAGGCTCTTACGCTGAGGCGGAGGCGGAGGCGGCGGCTCTCTCTC 960
Db 877 TCGAGTCTGCTGAGGCGGATGAGGCGGAGGCGGAGGCGGAGGCGGCGGCTCTCTCTC 936
QY 961 CCGTCTCTGAGGAGTCTGAGGCGGCGGCTGAGGCGGAGGCGGAGGCGGAGGCGGAGG 1020
Db 937 CCGTCTCTGAGGAGTCTGAGGCGGCGGCTGAGGCGGAGGCGGAGGCGGAGGCGGAGG 996
QY 1021 CCGAGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
Db 997 CCGAGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
QY 1081 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
Db 1057 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
QY 1141 GCGCGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 1117 GCGCGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
QY 1201 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
Db 1177 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
QY 1261 TTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
Db 1237 TTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
QY 1321 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
Db 1297 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1356
QY 1381 CCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Db 1357 CCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1416
QY 1441 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db 1417 GCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
|||||

RESULT	13
AR282639	
LOCUS	AR282639 2151 bp DNA linear PAT 10-APR-2003
DEFINITION	Sequence 9 from patent US 6521748.
ACCESSION	AR282639
VERSION	AR282639.1 GI:29719238
KEYWORDS	.
SOURCE	Unknown.
ORGANISM	Unknown. Unclassified.
REFERENCE	1 (bases 1 to 2151)
AUTHORS	Tang X.-S.
TITLE	Polynucleotide encoding a mutant Rhodotorula glutinis tyrosine ammonia lyase polypeptide
JOURNAL	Patent: US 6521748-A 9 18-FEB-2003;
FEATURES	Location/Qualifiers
source	1..2151 /organism="unknown" /mol_type="genomic DNA"
ORIGIN	
Query Match	65.7% Score 1626.4; DB 6; Length 2151;
Best Local Similarity	83.7%; Pred. No. 5e+163;

	Matches 1829;	Conservative	0;	Mismatches 322	Indels	35;	Gaps	5
QY	1	ATGGCCCCCTC	NTGATTCGATTCGGCACTCGTNTGCCAAGGCTC	TCNAAACGANTG	60			
Db	1	ATGGCACTCTGCTG	ACTTCGATCTGGCACTCTTTGGCAACGGGCTG	CTCCAAAG	60			
QY	61	CACGCGANNCC	GCNCCNNMNCNACGGGGCGCACGCTTCACNCTCGNCGCGCGNCG	120				
Db	61	CAGGCTGT-----	-----CAATGGGGCTTCAGCCAACTCT-----	GGAGTGGAG	100			
QY	121	GCTTCGCTCTCC	CCGACCACTCCAGNNACGAGCTCGACATCTGTTGAGNAGATCTCTGCGC	180				
Db	101	GCTCGCACTTCGCCA	CAACCCAGGTCCAGAGGTTCACATCTGTCAAGAAATGCTTCGCGC	160				
QY	181	ACCCCAACGNNAC	GNACGANNNTGAACTCGACGGGTACACCTCTCACCTCCGNNGACGTGC	240				
Db	161	CGGCCACGACCTCCA	---CGCTCGAACTCGACCGCTACTCTGCTCAACTCGGAGACGTGC	217				
QY	241	TCGGCGCGC	CNCGCAGGGGCGCGCNCTCCGCTCNACACAGNCGAGAGATCCGCGCA	300				
Db	218	TCGCGCGCCGAG	AAGGGACGGCTCTCCGCTCCAGAGACAG- GGAAGAGATCGCTCA	276				
QY	301	AAGATTCGACAA	NAGNTGAGTTCTCGNNCCACACTCNAAACAGAGTCTACGANTC	360				
Db	277	AAGATTGACAAAT	CGGTTCGATTTTCGCTCCGAACTCTCATAGCGGTCTTAGGGGCTC	336				
QY	361	ACGACTGTGTTT	CGCGGCTCGGCCGACACCCGGACTGAGGATGCAATCTCGCTCCAGAG	420				
Db	337	ACGACTGATTTGG	GGGATTCGACGACACCGCACCGAGAGAGGCACTCGCTCCAGAG	396				
QY	421	GCNCTCTTGAG	ACACACTCTGGGGTGTCTCCNACGTCGNTGANTCCTTCNCGCTC	480				
Db	397	GCCTCTCTGAG	ACAACACTCTGGGGTGTCTCCCTTGTGATCGATCTCGCTC	456				
QY	481	GGNCGGGGCTC	GAAATCGCTTCCGCTTGAGGTGTGCGGGCGCATGACATCGC	540				
Db	457	GGCGGGGTCTG	AGAACTCGCTTCCCTTGAGGTGTTGGGGGGCATGAAATCGC	516				
QY	541	GTCAACTGCTC	ACNCGGGCTACTGGGCGTTCGCTCTGTGCTCTGAGGCGCTCAC	600				
Db	517	GTCAACAGCTTGA	CCCCGGGCACTCGGCGTCCGGCTGTGATCTCGAAGCGCTCAC	576				
QY	601	AACTTCCTCAAC	CAAGGATCACCCCCATGTCCCCCTCGGAGCAACATCTCGGCGTC	660				
Db	577	AACTTCCTCAAC	AGAGATCACCCCATGTCCCCCTCGGGGACATCTTCGCGTC	636				
QY	661	GGGACCTCTCC	CCNCTCTNTACATGCGCGCGCATACCGGTTCACCCGACNNCAAG	720				
Db	637	GGGACCTCTCTC	CTCTCTCTACATTCAGCGGCATCAGGGCTCACCCGACAGCAAG	696				
QY	721	GTNCAAGTNNN	CAGAGGGGANNAGAAATGATNTNGCCCGGAGAGCATTCGCGTC	780				
Db	697	GTGCAAGTGTCA	AGAGGAGAGAAATCTTACGCCCCGAGGCGATGGCGCTC	756				
QY	781	TTNGGCTCGA	CCCGTTCCTTCGCGCCCAAGAGGATCTCGATCTGTCAACGACG	840				
Db	757	TTCAACTTCGA	CGCCGTGTCTTGCGCCGAAGAGGATCTCGATCTGTCAACGAC	816				
QY	841	GGCGTCTCGG	CTCGATGGGACCTTCGTCTTCGACAGAGCACATGCTCTCGCTCTC	900				
Db	817	GGCGTCTCAGA	TCATGAGCACTTCGTCTTCGACAGAGCACATGCTCTCGCTCTC	876				
QY	901	TGCGAGGCGCT	CACGAGCTTACAGGTTCGAGGCATGTCGACAGCGGCTGTGTTCCAC	960				
Db	877	TGCGAGTGGCT	CAAGGCAATGAGGTTCGACAGCGGCTGTGTTCCAC	936				
QY	961	CCNTTCTTCAG	AGATCGCGCTCTCACCCGACCAATGAGAGTGGCGCAACATC	1020				
Db	937	CCCTTCTCTTCA	CGAGTTCAGCGCCCTCACCCGACCAATGAGAGTGGCGGAAATCTC	996				
QY	1021	CGCAGGCTCTC	AGAGGAGAGNTTTCAGGTCTCAACAAGAGAGAGGTCAAGGTCAAG	1080				
Db	997	CGCAGGCTCTC	AGAGGAGCGCTTTGCTGTCTCAACAAGAGAGAGGTCAAGGTCAAG	1056				

1081 GACGACGAGGAGATTTCTCCGACGAGACCGCTACCGCTCCGACGTCAGTGGCTC 1140
1057 GAGAGAGAGGAGATTCTCCGACGAGACCGCTACCGCTCCGACGTCAGTGGCTC 1116
1141 GGGCCGCTGTGAGGAGATGATTAAGCCGACGCGCTCTCTGCTGAGGCGGACAG 1200
1117 GGGCCGCTGTGAGGAGATGATTAAGCCGACGCGCTCTCTGCTGAGGCGGACAG 1176
1201 TCGACGACGAGACGAGCCGCTGATGACGCTGAGAACGAGAACGACGAGGCGGAC 1280
1177 TCGACGACGAGACGAGCCGCTGATGACGCTGAGAACGAGAACGACGAGGCGGAC 1236
1261 TTCCAGGCGGCTGCTGCGACGAGACGATGAGAAAGACTGCGCTGCGCTGCGCTGATC 1320
1237 TTCCAGGCGGCTGCTGCGACGAGACGATGAGAAAGACTGCGCTGCGCTGCGCTGATC 1296
1321 GGGAGGCTGATCTGACGAGGCTGACGAGGCTGACGAGGCTGACGAGGCTGACGAG 1380
1297 GGGAGGCTGATCTGACGAGGCTGACGAGGCTGACGAGGCTGACGAGGCTGACGAG 1356
1381 CCNTCTGCTGCTGCTGCGACGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
1357 CCCTCTGCTGCTGCGGCGGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1416
1441 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
1417 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
1501 GAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
1477 GAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1536
1561 ACNCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
1537 ACNCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1596
1621 GAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1597 GAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1656
1681 NCNCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
1657 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1716
1741 CTGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
1717 CTGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1776
1801 CTGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
1777 CTGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1836
1861 NNCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
1837 TCGTTCGAGCTC-----GCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1885
1921 CCGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
1886 CCGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1945
1981 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
1946 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2005
2041 TCGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
2006 TCGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2065
2101 TCGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
2066 TCGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2125

QY 2161 ACGTCTCTGTCAGATGCTCGCNTAG 2186
Db 2126 ACGTCTCTGTCAGATGCTCGCNTAG 2151
RESULT 14
AX082544 2151 bp DNA linear PAT 28-FEB-2001
LOCUS AX082544
DEFINITION Sequence 9 from Patent WO0111071.
ACCESSION AX082544
VERSION AX082544.1 GI:13184657
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Tang, X.S., Vannelli, T.M., Qi, W.W., Sariastani, S. and Garenby, A.A.
TITLE Bioproduction of para-hydroxycinnamic acid
JOURNAL Patent: WO 011071-A 9 15-FEB-2001;
E.I. DU PONT DE NEVOURS & COMPANY INCORPORATED (US)
FEATURES
source
1..2151
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="mutant from Rhodotorula glutinis"
ORIGIN
Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
Best Local Similarity 83.7%; Pred. No. 5e-169;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;
QY 1 ATGGCCCTCTCCNTGACTCGATGCGACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
Db 1 ATGGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
QY 61 CAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db 61 CAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
QY 121 GCTGCTCTCTCCGACCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
Db 101 GCTGCGACCTGCGCCCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 160
QY 181 ACCCGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db 161 GCGCGACGAGCTGCA---CGCTGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 217
QY 241 TCGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 218 TCGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 276
QY 301 AAGATGACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
Db 277 AAGATTTGACAAATCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 336
QY 361 ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
Db 337 ACGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 396
QY 421 GCGTCTCTGAGACCAAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
Db 397 GCGTCTCTGAGACCAAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 456
QY 481 GCGTCTCTGAGACCAAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
Db 457 GCGTCTCTGAGACCAAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 516
QY 541 GTCAACTGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
Db 517 GTCAACTGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 576
QY 601 AACTTCTGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660

NRVAAESAISLTROVRETFWASAATSPPALSTYLSPTQLILVAFVREELGVKARQDV
ELGKEVITIGSNVSKIVYIAISGRINNVLKMLA"
1..2151
misc_feature /note="mutant from Rhodotorula glutinis"

Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
Best Local Similarity 83.7%; Pred. No. Se-169;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGGCCCCCTCCNTGACTCGATCGCGACCTTGANTCGCCAA CGGCTCNCNAA CGSANTNG 60
DB 1 ATGGACACCTCGCTCGACTCGATCTCGCACTCGTTCCGAAAGCGCTGCGCAAG 60
QY 61 CACGCGNNCCGNCNNNGCNCNACGGCGCGCAGTCCACNCTCNCNGCGCGCGNG 120
DB 61 CAGGCTGT-----CAATGCGCTCGACCAACCTC-----GCAATCGAG 100
QY 121 GCTCGCTCTCTCCCGACCAACCCAGNNAGACGACGTGACATGTGAGNAGATCTCTCGCG 180
DB 101 GCTCGCACTCTGCCCAACCCAGGTCAACGACAGTGCATCGTCGAGAAATGCTCGCGG 160
QY 181 ACCCCACCGNNAACNACGNNNTGAACTTCGAGGGTACACCTCACTCGGNAAGCTCG 240
DB 161 CGCGCACCGACTGA---CGCTCGAACTCGAGCGCTACTCGCTCAACTCGAGAGCGTGC 217
QY 241 TCGGCGCGCNCNCGAAGGCGCGCNGCTCGCGGTCCGATCAGACAGNCGACGAGATCGCGCA 300
DB 218 TCTCGCGCGCGAGGAAGGCGACGCTGTCCGCTCAAGACAG---GAGAGATCGCTCA 276
QY 301 AAGATCGACANAGNGTGAATTCTCCGANNCAAGCTCMAACAAGNGTCTAAGSNGTC 360
DB 277 AAGATTGACAAATCGGTGAGTTCGTGGCGCTCGCAACTCTCATGAGGCTGTACGGCGTC 336
QY 361 AGCAGCTGTTTGGGGGCGCTCGGCGCACCCGGAATGAGATGCAATCGCTCGCAGAG 420
DB 337 ACACCTGGAATTGGCGGATTCGGGACACCCGACCGAGAGCGCATTCCTCTCAAGAG 396
QY 421 GGNCTCTCGAGACCAAGCTCTGCGGTGTNCTCCNAGCTGANTCGANTCTTNGCCTC 480
DB 397 GCTCTCTCGAGAGACCAAGCTCTGCGGTGTNCTCCCTGCTGTTGATCTCGCTCGCCTC 456
QY 481 GGNCGGGCTCGAAGAACTCGCTTCGCTCGAGGTCTGTCGGGGGCAATACATCCGCG 540
DB 457 GGGCGGGCTCGAAGAACTCGCTTCCTCCCTCGAGGTGTTCGGGGGCATACATCCGCG 516
QY 541 GTCAACTCGCTCAACNCGCGCACTCGGCGTCCGCTCGTGTCTCGAGGCGCTCAAC 600
DB 517 GTCAAGAGCTTGAACCGCGCGCACTCGGCTGTCCGCTGTGTCTGTCTGAGAGCGCTCAC 576
QY 601 AACTTCTCAACCAAGGATCAACCCCATCTGTCCCTTCGCGCGACCATCTCGCGTGC 660
DB 577 AACTTCTCAACCAAGGATCAACCCCATCTGTCCCTTCGCGCGACCATCTCGCGTGC 636
QY 661 GGGGACCTCTCCGCTCTGNTAATCGCGCGCGCATCACCGGTCAACCGGACNNCAAG 720
DB 637 GGGGACCTCTCTCTCTCTCTCAATGAGCGGCGCATGAGGTACCCCGGACGCAAG 696
QY 721 GTNACGNTNNTNACAGAGGCAANNAGAAATCATGTNCGCCGCGAGCGANTCGCGCTC 780
DB 697 GTCAAGGTGTCAACGAGGGAAGAAATCTGTAGCCCGCGAGAGCGATGCGCGCTC 756
QY 781 TTNAGTCTGAGCGCGCTCTCTCGGCGCGAAGAGGGTCTGTGTCTCAAGCGACG 840
DB 757 TTNAACTTGAAGCGCGCTCTCTCGGCGCGAAGAGGGTCTGTGTCTCAAGCGACG 816
QY 841 GCGGCTCTCGGCTCGATGCGGACCTCGCTCTGACGACGACATGCTCTCGCTCTC 900
DB 817 GCGGCTCTGAGATCGATGCGGACCTCGCTCTGAGAGAGCAATGCTCTCGCTCTC 876
QY 901 TCGCAGGCGCTCAAGGCTCTMACGTCGAGGCGATGTCGACAGCGCGGCTCGTTCCAC 960
DB 877 TCGCAGGCGCTCAAGGCGCTCAAGGCGATGTCGAGGCGACAGCGCGGCTCGTTCCAC 936

QY 961 CQNTTCCTCCACGAGCTCAGCGGCGCTCAACCCGACCCAGATCGAGGTCTCGCGCAATC 1020
DB 937 CCTTCTCTTCAAGAGTCAAGGCTCAGCGGCGCTCAACCCGACCGAGATCGAAGTCTCGGGAAATC 996
QY 1021 CGACGCTCTCTGAGAGGCAAGGCTTCCGCTCCACCAAGAGAGAGGTCAAGTCAAG 1080
DB 997 CGCAAGCTCTCTGAGAGGAGCGGCTTGTCTGTCCACCATGAGAGAGGTCAAGTCAAG 1056
QY 1081 GACGACGAGGGCATTTCTCCGCAAGACCGCTACCCGCTCCGCAAGTCCGCTCAGTGGCTC 1140
DB 1057 GACGAGAGGGCATTTCTCCGCAAGACCGCTACCCGCTCCGCAAGTCTCTCAAGTGGCTC 1116
QY 1141 GGGCGCTCTCTGAGGACATGATTCAGCCCAAGGCTCTCTCTCTGAGGCGGNCAG 1200
DB 1117 GGGCGCTCTCTGAGGACATGATTCAGCCCAAGGCTCTCTCTCTGAGGCGGNCAG 1176
QY 1201 TCGACGACCGACCAACCGCTCATGAGAGTCAAGAAACAGANAGCCCAACGAGGCGAAC 1260
DB 1177 TCGACGACCGACCAACCGCTCATGAGAGTCAAGAAACAGATTCGACCAAGGCGGAAT 1236
QY 1261 TTCGAGGCGNCGCTGTGCGNAAACAGATGAGAGAGACTCGCTCGGCTCGGCTGATC 1320
DB 1237 TTCGAGGCTGCGCTGTGCGCAACCATGAGAGAGACTCGCTCGGCTGCGGCTGATC 1296
QY 1321 GGCAGGCTCAACTTCAGCGAGCTCAACGAGTCTCAACGCGGCGATGAACCGGCGCTM 1380
DB 1297 GGCAGGCTCAACTTCAGCGAGCTCAACGAGTCTCAACGCGGCGATGAACCGGCGCTC 1356
QY 1381 CGNTCTGCTCTGCGCGGAGAGACCCGCTCGCTCATATCTGCAAGGCGCTGACATT 1440
DB 1357 CCTCTCTGCTCTGCGCGGAGAGACCCGCTCGCTCTCTCAACCTGCAAGGCGCTGACATC 1416
QY 1441 GCGCGGNGGNCNTACATTCGAGAGTCTGAGACCTTGCAACCGGATNACGACCACGTC 1500
DB 1417 GCGCGTGGGGGTACACTCGGAGTGGGACACTCGCAACCTTGAGAGAGCAAGTC 1476
QY 1501 CAGCGGCGNAGATGGGCAACCAAGGCGTCAACTCTGCTCGNCTCATCTCGGCGCGCGC 1560
DB 1477 CAGCGGCTGAGATGGCAACCAAGGCGTCAACTCTGCTCGNCTCATCTCGGCTGCTGTCG 1536
QY 1561 ACGGCGGAGGCAACGAGTCTCTCTCTCTCTCGGCAACCACTCTAATGAGTCTC 1620
DB 1537 ACGGCGGAGTCAAGAGTCT 1596
QY 1621 CAGGCGTCTGACTCTCGCGGAGTGGAGTTGCAAGTTCAAGAAAGCTTGACCCGNTNNTC 1680
DB 1597 CAGGCGATGACTTTCGCGGCGACCGAGTTGAGTTCAAGAAAGCTTGACCCGACATC 1656
QY 1681 NCMNGCTCTNTNAGACACTTGTGCGNCGCCCTCGACGCGNNNNNNNNNNNAAGAA 1740
DB 1657 GTCTGCTATGACACAGACTTGTGCTCGCATGACCGGCTGGAACCTGCGGAGAG 1716
QY 1741 CTGAGGACCAAGTCAACAGNCGCTCNACAAAGGCTTGAAGACCACTCTGAC 1800
DB 1717 CTGCTGAGAAAGTGAACAAAGCGCTCGCAAGGCGCTGAGACCACTCTGAC 1776
QY 1801 CTGAGCGCGCTGAGCAAGAGCTTCTGCTTCTGAGACCGGCACTGCTGAGANNNNNN 1860
DB 1777 CTGCTCGCGCTGAGCAAGAGCTTCTGCTTCTGCGCGCGGACCGTGTGAAGTCTC 1836
QY 1861 NNCTCTGCGCNNNNGCCANNAGTCTGCTCGGAGCGGTCAAGCTGGAAGTCTGCT 1920
DB 1837 TCGTGAAGTCT-----GCTCTGCTGCTGCGCGCTCAAGCTGGAAGTCTGCTG 1885
QY 1921 CGGCGGAGAGGCACTTCTGCTCAACGCGCANAGTCTCGGACNCTTCTGAGGCGNCCGT 1980
DB 1886 CGGCGGAGGCGCACTTCTGCTCAACGCGCAAGTCTCGGACNCTTCTGAGGCGNCCGT 1945
QY 1981 CGTCTGCTGCGCGCGGCTCNGTCACTCTGCGCGGCGACGCGCTCTGACTGCTTTCG 2040
DB 1946 CGACTCTGTCGCGCGGCTCTGTACTCTTCTGCGCGGCACTCAAGTCTCTAAGCTTTCG 2005

```
QY 2041 TCCGCGAGAGCTCGGCTCAAGGCCCGCGCGCGAGCTTTCTCTCGGCAAGCAGAGG 2100
Db 2006 TCCGCGAGAGCTTGGCGTCAAGGCCCGCGCGAGAGCTTTCTCTCGGCAAGCAGAGG 2065
QY 2101 TGACGATCGGCGACCAACGCTCTCCGCACTCTACGAGGCCATCAAGNNCGGCGNGCATCAAC 2160
Db 2066 TGACGATCGGCTCGAAGCTCTCCAGATCTACGAGGCCATCAAGTGGGCAAGATCAACA 2125
QY 2161 ACGTCTCGTCAAGATGCTCGCTAG 2186
Db 2126 ACGTCTCTCTCAAGATGCTCGCTAG 2151
```

Search completed: September 11, 2004, 23:12:51
Job time : 15249.4 secs


```
QY 1861 NNCTCTCGCCNNNNCCANNAGTCTCGCTCGCGCCGCTCAACGCTGAGAGTGGCT 1920
DB 1861 NNCTCTCGCCNNNNCCANNAGTCTCGCTCGCGCCGCTCAACGCTGAGAGTGGCT 1920
QY 1921 CCGCGAGAGAGCTCTCTGCTCAACGCGCANNAGTCCGGAACCTCTTGCGCGCCGCT 1980
DB 1921 CCGCGAGAGAGCTCTCTGCTCAACGCGCANNAGTCCGGAACCTCTTGCGCGCCGCT 1980
QY 1981 CGTCTCTGCTCGCCGCGCTCGCTCTGCTCGCGCGCGAGCGGCTCTGCTGAGCTGG 2040
DB 1981 CGTCTCTGCTCGCCGCGCTCGCTCTGCTCGCGCGCGAGCGGCTCTGCTGAGCTGG 2040
QY 2041 TCCGCGAGAGAGCTCTGCTCAACGCGCGCGCGCGAGCTCTCTGCTGAGAGAGG 2100
DB 2041 TCCGCGAGAGAGCTCTGCTCAACGCGCGCGCGCGAGCTCTCTGCTGAGAGAGG 2100
QY 2101 TACGATCGGACCAAGCTCTTCCGCACTCAAGAGCATTAAGNCGCGCATCAAC 2160
DB 2101 TACGATCGGACCAAGCTCTTCCGCACTCAAGAGCATTAAGNCGCGCATCAAC 2160
QY 2161 ACGTCTCTGCTCAAGATGCTCGCTAGNNCCNNCCANNCTCGCTNNNNCCNNCC 2220
DB 2161 ACGTCTCTGCTCAAGATGCTCGCTAGNNCCNNCCANNCTCGCTNNNNCCNNCC 2220
QY 2221 NNCCNNNNNNNNCTNNNNNNNNCTGANTCTGTCNNNNNNCGANNNNNNNN 2280
DB 2221 NNCCNNNNNNNNCTNNNNNNNNCTGANTCTGTCNNNNNNCGANNNNNNNN 2280
QY 2281 TNNNNCTNNCTNNCTCNCNNNNCCNNCTGNTCTNNNNNNCTNNNNNNNNNN 2340
DB 2281 TNNNNCTNNCTNNCTCNCNNNNCCNNCTGNTCTNNNNNNCTNNNNNNNNNN 2340
QY 2341 NCTNNCCANNACNNCTNNNNNNNNCCANNCCANNCCANNCCANNCCANN 2400
DB 2341 NCTNNCCANNACNNCTNNNNNNNNCCANNCCANNCCANNCCANNCCANN 2400
QY 2401 NNNNAAA 2408
DB 2401 NNNNAAA 2408

RESULT 2
AAD45816
ID AAD45816 standard; DNA; 2475 BP.
XX AAD45816;
AC AAD45816;
XX 07-AUG-2003 (revised)
DT 27-DEC-2002 (first entry)
XX
DE Phenylalanine ammonia lyase consensus DNA.
XX
KW Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytostatic; virucide; ds.
XX
OS Rhodotorula glutinis.
OS Rhodotorula mucilaginosa.
OS Rhodospiridium toruloides.
XX
PN US2002102712-A1.
XX
PD 01-AUG-2002.
XX
PF 24-AUG-2001; 2001US-00939408.
XX
PR 24-JUL-2000; 2000US-00624693.
XX
PR 24-JUL-2001; 2001WO-US023270.
XX
PA (PCBU-) PCBU SERVICES INC.
XX
XX Yoshida RK, Kootstra AB;
PI
```

```
XX
DB MPI; 2002-690616/74.
XX
PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT human immunodeficiency virus or human cytomegalovirus infection.
XX
PS Claim 5; Page 45-46; 74pp; English.
XX
CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC such proteins. PAL sequences are useful for producing L-phenylalanine.
CC They are useful for treating mammals having diseases, disorders or
CC conditions that would benefit from treatment with PAL proteins such as
CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC human cytomegalovirus (hCMV) infection. Sequences of the invention are
CC also used in gene therapy. The present sequence is PAL consensus DNA.
CC (Updated on 07-AUG-2003 to correct OS field.)
XX
SQ Sequence 2475 BP; 350 A; 869 C; 568 G; 341 T; 0 U; 347 Other;
XX
Query Match 86.0%; Score 2126; DB 6; Length 2475;
Best Local Similarity 100.0%; Pred. No. 2.6e-268;
Matches 2408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCCCCCTTCCTGATCGATCGGACCTCTGTCGCAACGCGCTCNCNACGNTG 60
DB 1 ATGCCCCCTTCCTGATCGATCGGACCTCTGTCGCAACGCGCTCNCNACGNTG 60
QY 61 CACGCGGNNCCGNNNNCCNNCCGNNCCGAGGCGCACGCTCCACGCTCNCNCCGCGCG 120
DB 61 CACGCGGNNCCGNNNNCCNNCCGNNCCGAGGCGCACGCTCCACGCTCNCNCCGCGCG 120
QY 121 GCTGCTCTCTCCGACCAACCCAGNAGAGAGCTCGACATCGTNGAGNAGATCCTG 180
DB 121 GCTGCTCTCTCCGACCAACCCAGNAGAGAGCTCGACATCGTNGAGNAGATCCTG 180
QY 121 GCTGCTCTCTCCGACCAACCCAGNAGAGAGCTCGACATCGTNGAGNAGATCCTG 180
DB 121 GCTGCTCTCTCCGACCAACCCAGNAGAGAGCTCGACATCGTNGAGNAGATCCTG 180
QY 181 ACCCCACGNNACGNNAGNNNTGAACTCGAGCGGTACACCTTCACCTCTGNGAGCTG 240
DB 181 ACCCCACGNNACGNNAGNNNTGAACTCGAGCGGTACACCTTCACCTCTGNGAGCTG 240
QY 181 ACCCCACGNNACGNNAGNNNTGAACTCGAGCGGTACACCTTCACCTCTGNGAGCTG 240
DB 181 ACCCCACGNNACGNNAGNNNTGAACTCGAGCGGTACACCTTCACCTCTGNGAGCTG 240
QY 241 TCGGCGCGCGCCGAGAGGCGCGCCGCTCNCAGACAGNCGACGAGATCCGCGCA 300
DB 241 TCGGCGCGCGCCGAGAGGCGCGCCGCTCNCAGACAGNCGACGAGATCCGCGCA 300
QY 301 AAGATCGACAAAGAGTGAATCTCTCCGNNCCAGCTCNAACAAGNGTCTACGNGTC 360
DB 301 AAGATCGACAAAGAGTGAATCTCTCCGNNCCAGCTCNAACAAGNGTCTACGNGTC 360
QY 361 ACGACTGGTTTCGGCGGCTCGGCGGACCCGAGCTAGAGATGMACTCTGCTCGAAG 420
DB 361 ACGACTGGTTTCGGCGGCTCGGCGGACCCGAGCTAGAGATGMACTCTGCTCGAAG 420
QY 421 GNCCTCTCGAGCACCAAGCTCTGCGGTGTCCTCCNACGTCGNTCGANTCTCNGCTC 480
DB 421 GNCCTCTCGAGCACCAAGCTCTGCGGTGTCCTCCNACGTCGNTCGANTCTCNGCTC 480
QY 481 GANCGGCGCTCGAGAACTGCTTCCGCTGAGAGTCTCGCGGCGCGCATGACATCCG 540
DB 481 GANCGGCGCTCGAGAACTGCTTCCGCTGAGAGTCTCGCGGCGCGCATGACATCCG 540
QY 541 GTCAACTGCTCAGNCGGCGGCACTCGGCGCTCGCTCGCTCGAGGCGCTCAC 600
DB 541 GTCAACTGCTCAGNCGGCGGCACTCGGCGCTCGCTCGCTCGAGGCGCTCAC 600
QY 601 AACTTCTCTCAACGAGGATCACCCCATGTCCTCCGCGGACCATCTCGGCGTGG 660
DB 601 AACTTCTCTCAACGAGGATCACCCCATGTCCTCCGCGGACCATCTCGGCGTGG 660
QY 661 GCGGACCTCTCCGCTCTGATGATGCGCGCGCATCACCGGTCACCCGAGCANNCAAG 720
DB 661 GCGGACCTCTCCGCTCTGATGATGCGCGCGCATCACCGGTCACCCGAGCANNCAAG 720
```


[illegible]

Db	1801	CTCGAGCGCGCGTGGCAAGACGCTTCTCGTTGCGCACCGGACCGTGTCTGAGNNNNN	1860
Qy	1861	NNGTCTTGCCNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAAACGCTTGAAAGTTCGCT	1920
Db	1861	NNGTCTTGCCNNNNNGCCANNAGGTCTCGCTCGCNGCCGTCAAACGCTTGAAAGTTCGCT	1920
Qy	1921	CCGCCGAGAAGGCATCTTCGCTCAAGCGCANATTCGCGACACNCCTTTGAGCGCNCCT	1980
Db	1921	CCGCCGAGAAGGCATCTTCGCTCAAGCGCANATTCGCGACACNCCTTTGAGCGCNCCT	1980
Qy	1981	CGTGTCTGCTGCGCGCGCTCNCGTACTCTTCGCGCGACGCGCGCTGTCTGACTCGTTG	2040
Db	1981	CGTGTCTGCTGCGCGCGCTCNCGTACTCTTCGCGCGACGCGCGCTGTCTGACTCGTTG	2040
Qy	2041	TTCGGGAGAGTGGGGCTCAAGGCCGCGCGCGCGAGCGTCTCTCGCGCAAGCAGAGG	2100
Db	2041	TTCGGGAGAGTGGGGCTCAAGGCCGCGCGCGCGAGCGTCTCTCGCGCAAGCAGAGG	2100
Qy	2101	TGACGATCGGCACCAACGCTCTCCGCACTACGAGCCATCAAGNNCGCGCATCAAC	2160
Db	2101	TGACGATCGGCACCAACGCTCTCCGCACTACGAGCCATCAAGNNCGCGCATCAAC	2160
Qy	2161	ACGTCCTCGTCAAGATGCTCGCTAGNNNCNNNNNCNNNNCTCGCTNNNNNCNNNNC	2220
Db	2161	ACGTCCTCGTCAAGATGCTCGCTAGNNNCNNNNNCNNNNCTCGCTNNNNNCNNNNC	2220
Qy	2221	NNNNCNNNNNNCTTTGNNNTCGNCTCGTGCNNNNCGGANNNTNCCNNNNNN	2280
Db	2221	NNNNCNNNNNNCTTTGNNNTCGNCTCGTGCNNNNCGGANNNTNCCNNNNNN	2280
Qy	2281	TNNNCTNNCTNNCTCNCNNNNACNNGTCNNTTNNNNCTNNNNNTNNNNNCNNNC	2340
Db	2281	TNNNCTNNCTNNCTCNCNNNNACNNGTCNNTTNNNNCTNNNNNTNNNNNCNNNC	2340
Qy	2341	NGNNNCAANNACNTNNNNNNANNCGGANNGANTNANGNTNCGNNNCNNNNN	2400
Db	2341	NGNNNCAANNACNTNNNNNNANNCGGANNGANTNANGNTNCGNNNCNNNNN	2400
Qy	2401	NNNAAAAA 2408	
Db	2401	NNNAAAAA 2408	
RESULT 3			
AAD45824 standard; DNA; 2163 BP.			
XX	AAD45824		
XX	AAD45824;		
XX			
DT	27-Dec-2002	(first entry)	
XX			
DE	Yeast phenylalanine ammonia lyase DNA #2.		
XX			
XX	Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;		
KW	cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;		
KM	human cytomegalovirus infection; cytostatic; virucide; ds.		
OS	Rhodotorula graminis.		
XX			
XX			
Key	Location/Qualifiers		
FT	1..2163		
FT	/*tag= a		
FT	/product= "PAL protein"		
FT	/EC_number= "4.3.1.5"		
FT	/transl_except= (pos:13..15, aa:Xaa)		
FT	/note= "Xaa corresponds to Val, Leu, Phe"		
FT	/transl_except= (pos:34..36, aa:Xaa)		
FT	/note= "Xaa corresponds to Val, Leu, Phe"		
FT	/transl_except= (pos:46..48, aa:Xaa)		
FT	/note= "Xaa corresponds to Val, Leu, Phe"		
FT	/transl_except= (pos:49..51, aa:Xaa)		
FT	/note= "Xaa corresponds to Thr, Ala, Ser"		

FT /transl_except= (pos:55. .57, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:58. .60, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Leu or an in-frame stop
codon"
FT /transl_except= (pos:73. .75, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:79. .81, aa:Xaa)
FT /note= "Xaa corresponds to Pro, Ser"
FT /transl_except= (pos:82. .84, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro"
FT /transl_except= (pos:100. .102, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:106. .108, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:115. .117, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT /transl_except= (pos:118. .120, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Gly, Trp"
FT /transl_except= (pos:142. .144, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Thr, Met, Glu, Ala, Val,
Gln, Pro, Leu"
FT /transl_except= (pos:160. .162, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:166. .168, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Glu, Gln"
FT /transl_except= (pos:193. .195, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp, Val"
FT /transl_except= (pos:196. .198, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Leu"
FT /transl_except= (pos:226. .228, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:259. .261, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:277. .279, aa:Xaa)
FT /note= "Xaa corresponds to Asp, Ala"
FT /transl_except= (pos:304. .306, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Asn"
FT /transl_except= (pos:307. .309, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:325. .327, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:334. .336, aa:Xaa)
FT /note= "Xaa corresponds to Asp, His, Tyr"
FT /transl_except= (pos:340. .342, aa:Xaa)
FT /note= "Xaa corresponds to Arg, Ser"
FT /transl_except= (pos:349. .351, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:442. .444, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:448. .450, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:457. .459, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Phe"
FT /transl_except= (pos:460. .462, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp"
FT /transl_except= (pos:469. .471, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Gly, Arg"
FT /transl_except= (pos:475. .477, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:547. .549, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:667. .669, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:673. .675, aa:Xaa)
FT /note= "Xaa corresponds to Ser"
FT /transl_except= (pos:709. .711, aa:Xaa)
FT /note= "Xaa corresponds to Ser, Thr, Ile, Gly, Ala, Val,
Cys, Phe"
FT /transl_except= (pos:715. .717, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:721. .723, aa:Xaa)
FT /note= "Xaa corresponds to Val"

FT /transl_except= (pos:724. .726, aa:Xaa)
FT /note= "Xaa corresponds to Val, Leu, Phe"
FT /transl_except= (pos:735. .738, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Asn, Thr"
FT /transl_except= (pos:751. .753, aa:Xaa)
FT /note= "Xaa corresponds to Tyr, Ser, Phe"
FT /transl_except= (pos:775. .777, aa:Xaa)
FT /note= "Xaa corresponds to Leu, Phe"
FT /transl_except= (pos:913. .915, aa:Xaa)
FT /note= "Xaa corresponds to Leu"
FT /transl_except= (pos:955. .957, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1036. .1038, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Arg, Thr, Gln, Pro, Trp,
Ser or a stop codon"
FT /transl_except= (pos:1231. .1233, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Thr, Met"
FT /transl_except= (pos:1261. .1263, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1369. .1371, aa:Xaa)
FT /note= "Xaa corresponds to Leu"
FT /transl_except= (pos:1372. .1374, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1396. .1398, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1459. .1461, aa:Xaa)
FT /note= "Xaa corresponds to Gly"
FT /transl_except= (pos:1477. .1479, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:1498. .1500, aa:Arg)
FT /transl_except= (pos:1552. .1554, aa:Xaa)
FT /note= "Xaa corresponds to Thr"
FT /transl_except= (pos:1606. .1608, aa:Xaa)
FT /note= "Xaa corresponds to Val"
FT /transl_except= (pos:1666. .1668, aa:Xaa)
FT /note= "Xaa corresponds to Met, Ile, Val,
Leu"
FT /transl_except= (pos:1669. .1671, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Val, Leu"
FT /transl_except= (pos:1672. .1674, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Pro"
FT /transl_except= (pos:1675. .1677, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1681. .1683, aa:Xaa)
FT /note= "Xaa corresponds to Ile, Leu, Phe"
FT /transl_except= (pos:1684. .1686, aa:Xaa)
FT /note= "Xaa corresponds to Lys, Glu, Gln"
FT /transl_except= (pos:1699. .1701, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1714. .1716, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Tyr"
FT /transl_except= (pos:1723. .1725, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Ala, Val"
FT /transl_except= (pos:1741. .1743, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:1747. .1749, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Asp, Tyr"
FT /transl_except= (pos:1849. .1851, aa:Xaa)
FT /note= "Xaa corresponds to Pro"
FT /transl_except= (pos:1852. .1854, aa:Xaa)
FT /note= "Xaa corresponds to Pro, Ser"
FT /transl_except= (pos:1927. .1929, aa:Xaa)
FT /note= "Xaa corresponds to Glu, Asp, Gln, His, Tyr or a
stop codon"
FT /transl_except= (pos:1939. .1941, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Pro, Ser"
FT /transl_except= (pos:1948. .1950, aa:Xaa)
FT /note= "Xaa corresponds to Ala, Pro, Ser"
FT /transl_except= (pos:1978. .1980, aa:Xaa)
FT /note= "Xaa corresponds to Thr, Ala, Ser"
FT /transl_except= (pos:2122. .2124, aa:Xaa)
FT /note= "Xaa corresponds to Asn, Ser, Thr, Asp, Gly, Ala,
Tyr, Cys"
FT /transl_except= (pos:2128. .2130, aa:Xaa)

/note= "Xaa corresponds to Ser, Arg, Cys"

FT US2002102712-A1.
 XX 01-AUG-2002.
 PD 24-AUG-2001; 2001US-00939408.
 XX 24-JUL-2000; 2000US-00624693.
 PR 24-JUL-2001; 2001MO-US023270.
 XX (PCBU-) PCB SERVICES INC.
 XX Yoshida RK, Koetsura AB;
 PI WPI; 2002-690616/74.
 DR P-PSDB; AAE27943.
 XX
 PS Claim 7; Page 53-56; 74pp; English.
 CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC
 Query Match 78.5%; Score 1944; DB 6; Length 2163;
 Best local Similarity 93.2%; Pred. No. 1.2e-262;
 Matches 2037; Conservative 0; Mismatches 126; Indels 23; Gaps 7;

QY 1 ATGGCCCCCTCCNTGACTGATCGCGACCTCGANTCGCAACGGGCTGNCNAACGANTG 60
 DB 1 ATGGCCCCCTCCNTGACTGATCGCGACCTCGANTCGCAACGGGCTGNCNAACGANTG 60
 QY 61 CAGCGGNNCCGNNNGNCGNNAAGGCGCGCACTCCAGCTCCGNGCGCGCGCGCG 120
 DB 61 CAGCGGCTCCGHCNAAGTGGSCMAAGGCGCGCACTCCAGCTCCGNGCGCGCGCGCG 120
 QY 61 CAGCGGCTCCGHCNAAGTGGSCMAAGGCGCGCACTCCAGCTCCGNGCGCGCGCGCG 119
 DB 61 CAGCGGCTCCGHCNAAGTGGSCMAAGGCGCGCACTCCAGCTCCGNGCGCGCGCGCG 119
 QY 121 GCTCGCTCTCCGACCAACCCAGNNAAGCAGCTGACATCGTNGAGNAGTCTCGCG 180
 DB 120 GCTCGCTCTCCGCG-CCACCAGVHAGCAGACTGACATCGTNGAGNAGTCTCGCGCG 178
 QY 121 GCTCGCTCTCCGACCAACCCAGNNAAGCAGCTGACATCGTNGAGNAGTCTCGCG 180
 DB 120 GCTCGCTCTCCGCG-CCACCAGVHAGCAGACTGACATCGTNGAGNAGTCTCGCGCG 178
 QY 181 ACCCCACCGNNAAGNNTGAACTCGAAGGGTACACCTCAACCTCGGNAAGCTCG 240
 DB 179 ACCCCACCGAC--GACGWSVTCGAATCGAGGGTACACCTCAACCTCGGNAAGCTCG 235
 QY 241 TCGGCGCGCCNCGCAAGGCGCGCNCNGTCCGCTGNCAGACAGNCGAGNATCCGCGCA 300
 DB 236 TCGGCGCGCCBCCGCAAGGCGCGCHBGTCCGCTC-CAGACAGNCGAGNATCCGCGCA 294
 QY 301 AAGATGACAAANAGNCTGAGTTCTCCGANNCCAGCTCCNACAAAGNCTTACGNGTC 360
 DB 295 AAGATGACAAANAGNCTGAGTTCTCCGANNCCAGCTCCNACAAAGNCTTACGNGTC 354
 QY 361 ACGACTGTTTGGCGGCTCGGCGCAGACCCGGAATGAGATGAGNATCGCTCCAGAG 420
 DB 355 ACGACTGTTTGGCGGCTCGGCGCAGACCCGGAATGAGATGAGNATCGCTCCAGAG 414
 QY 421 GCNCTCTCGAGACCAAGCTTGCGGTGTCCTCCNAGCTGANTGANTCTTCNGCTTC 480
 DB 415 GCGCTCTCGAGACCAAGCTTGCGGTGTCCTCCNAGCTGANTGANTCTTCNGCTTC 474
 QY 481 GCGGCGGCTCGAGAACTCGCTCGCTCGAGGCTGCTCGGCGGCGCATGACATCCGC 540
 DB 475 GCGGCGGCTCGAGAACTCGCTCGCTCGAGGCTGCTCGGCGGCGCATGACATCCGC 534
 QY 541 GTCACCTGCTCAACNCGCGGCACTCGGCGNCTCGCTCGTGTCTCGAGGCGCTCAC 600
 DB 535 GTCACCTGCTCAACNCGCGGCACTCGGCGNCTCGCTCGTGTCTCGAGGCGCTCAC 594
 QY 601 AACTTCTCAACCAAGCATCAACCCCATGTCCTCCCTCGCGGAGCATATCGGCGTGC 660

DB 595 AACTTCTCAACCAAGCATCAACCCCATGTCCTCCCTCGCGGAGCATATCGGCGTGC 654
 QY 661 GCGGACCTCTCCGNCCTCTGNTATACGCGCGCCATCACCGGTACCCGAGCANNK 720
 DB 655 GCGGACCTCTCCGNCCTCTGNTATACGCGCGCCATCACCGGTACCCGAGCDBCAAG 714
 QY 721 GTNCACTGNTTCAAGAGGCGCANNAGAGATATGTCGCGCGGAGGCGATCGCTTC 780
 DB 715 GTBACGTYKTSCACAGAGGCGCAGAGAGATATGTCGCGCGGAGGCGATCGCGCTTC 774
 QY 781 TTNGTCTCGAGCCCGCTGCTCGGCGGAGAGGAGTGGGTCTCGTCAACGCGACG 840
 DB 775 TTGGTCTCGAGCCCGCTGCTCGGCGGAGAGGAGTGGGTCTCGTCAACGCGACG 834
 QY 841 GCGGTCTCGGCTCGATCGGCAACCTCGCTCTGCAAGCGACACATGCTTGTCTCTC 900
 DB 835 GCGGTCTCGGCTCGATCGGCAACCTCGCTCTGCAAGCGACACATGCTTGTCTCTC 894
 QY 901 TCGCAGGCGCTCAAGGCTCTTACGCTGAGGCGCATGTCGCGCACCGCGCTGTCGAC 960
 DB 895 TCGCAGGCGCTCAAGGCTCTTACGCTGAGGCGCATGTCGCGCACCGCGCTGTCGAC 954
 QY 961 CCNTTCTCGACGAGCTCAAGCGGCTCAACCGACCGAGTCGAGTTCGCGCAACATC 1020
 DB 955 CCNTTCTCGACGAGCTCAAGCGGCTCAACCGACCGAGTCGAGTTCGCGCAACATC 1014
 QY 1021 CGCAGCTCTCGAGGCGAGCNGTTTCCGTCACACGAGAGAGAGTCAAAGTCAAG 1080
 DB 1015 CGCAGCTCTCGAGGCGAGGCTTTCGCTCACACGAGAGAGAGTCAAAGTCAAG 1074
 QY 1061 GACGACGAGGAGCATTCCTCCGCGAGGAGCGCTACCGCTCGGACGTCGCTGATGTC 1140
 DB 1075 GACGACGAGGAGCATTCCTCCGCGAGGAGCGCTACCGCTCGGACGTCGCTGATGTC 1134
 QY 1141 GCGCGCTCTCGACGAGCATGATTCACGCGCAGCGNGTCTCGTCTGAGGCGGAGCAG 1200
 DB 1135 GCGCGCTCTCGACGAGCATGATTCACGCGCAGCGNGTCTCGTCTGAGGCGGAGCAG 1191
 QY 1201 TCGAGACCGACCAACCGGCTCTACGAGCTGAGAGCAAGNAGCCACACCGCGGCAAC 1260
 DB 1192 TCGAGACCGACCAACCGGCTCTACGAGCTGAGAGCAAGNAGCCACACCGGCGCAAC 1251
 QY 1261 TTCAGCGGNCGCGCTGTCGNAACAGATGAGAGACTCGCTCGCNCCTCGCTGATC 1320
 DB 1252 TTCAGCGGNCGCGCTGTCGNAACAGATGAGAGACTCGCTCGCNCCTCGCTGATC 1311
 QY 1321 GGCAGGCTCAACTTACGAGCTTCAACGAGATGTCACGCGGATGAACCGGCGCTTN 1380
 DB 1312 GGCAGGCTCAACTTACGAGCTTCAACGAGATGTCACGCGGATGAACCGGCGCTTN 1371
 QY 1381 CCNTTCTCGACGAGCTCAACGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 DB 1372 CCNTTCTCGACGAGCTCAACGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431
 QY 1441 GCGCGGNCGNTACACTTGGAGCTGAGNACCTTGCCAACTCGGTNAGACCTCACGTC 1500
 DB 1432 GCGCGGNCGNTACACTTGGAGCTGAGNACCTTGCCAACTCGGTNAGACCTCACGTC 1491
 QY 1501 CAGCGCGNAGATGAGGAGAACAGGCGGTCAACGCTGCGNCTCATCTCGGNCGCGCG 1560
 DB 1492 CAGCGCGNAGATGAGGAGAACAGGCGGTCAACGCTGCGNCTCATCTCGGNCGCGCG 1551
 QY 1561 ACNCGGAGGCAACGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
 DB 1552 ACNCGGAGGCAACGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1611
 QY 1621 CAGGCGCTGACCTTCGCGCGATGAGTTCGAGTTCAAGAGAGATTCGACCGGNNTC 1680
 DB 1612 CAGGCGCTGACCTTCGCGCGATGAGTTCGAGTTCAAGAGAGATTCGACCGGNNTC 1671
 QY 1681 NCNNCGCTCMTNAGACGACACTTGGCNGCGCTCGACGCGCNCNNNNNNNNNNNACGA 1740

```

Db      1672 VCBDCGCTGTCVAGCAGCACTTTGGCDYGCCTCGACGGC-----WAGGAA 1719
QY      1741 CTGGNGGACAAGGTCAACAGAGCGCTCNAACAGCGNCTCGAGGACCACTGCTTAAGAC 1800
Db      1720 CTGGAGGACAAGGTCAACAAGGCTCDAACAAGCGBTGAGAGACCACTGCTTAAGAC 1779
QY      1801 CTGAGACCGCGCTGGGACGACGCTTCTGTTGCGACCGGACCGCTGTCGAGNNNNN 1860
Db      1780 CTGAGACCGCGCTGGGACGACGCTTCTGTTGCGACCGGACCGCTGTCGAGCTCTC 1839
QY      1861 NGGTCTCTCCGNNNNGCCANNAGGTCTGCTTGCNCGCGCTCAAGCGCTGGAAGTGGCT 1920
Db      1840 TGTGCTCTCCGCTGCTGCA--AGGTCTGCTGCGGCGGTAAAGCTGGAAGTGGCT 1897
QY      1921 CCGCCGAGAGGCACTCTGCTCAAGCGGNNANGTCCGAGNCCCTTTCGAGCGGCGCT 1980
Db      1898 CCGCCGAGAGGCACTCTGCTCAAGCGGNNANGTCCGAGNCCCTTTCGAGCGGCGCT 1957
QY      1981 CGTCTGCTGCGCCGCGCTCNCGTAACCTCTGCGCGGCAAGCGGCTGCTGCTTGC 2040
Db      1958 CGTCTGCTGCGCCGCGCTCNCGTAACCTCTGCGCGGCAAGCGGCTGCTGCTTGC 2017
QY      2041 TCCGCGAGAGAGCTCGGCGCTCAAGGCGCGCGCGGCGAGAGTCTTCTGCGGCAAGAGAG 2100
Db      2018 TCCGCGAGAGAGCTCGGCGCTCAAGGCGCGCGCGGCGAGAGTCTTCTGCGGCAAGAGAG 2077
QY      2101 TGAAGATCGGCAACAAGTCTTCCGATCTACGAGCGCATCAAGNNCGGCGNCGATCAAC 2160
Db      2078 TACAGATCGGCAACAAGTCTTCCGATCTACGAGCGCATCAADPVGCGGCGATCAACC 2137
QY      2161 AGCTCTCTCTCAAGATGCTCGGCTNG 2186
Db      2138 AGCTCTCTCTCAAGATGCTCGGCTAG 2163

RESULT 4
ABA95243
ID      ABA95243 standard; cDNA; 2439 BP.
AC      ABA95243;
XX      07-AUG-2003 (revised)
DT      10-JUN-2002 (first entry)
XX      XX
DE      R. toruloides PAL polypeptide encoding cDNA.
XX      PAL; yeast; phenylalanine ammonia lyase; cinamic acid; cinamate; ss;
XX      cytotaxic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
OS      Rhodospiridium toruloides.
XX      XX
FH      Key Location/Qualifiers
FT      CDS 1..2151
FT      CDS /product= "phenylalanine ammonia lyase"
XX      XX
PN      MO200208402-A2.
XX      XX
PD      31-JAN-2002.
XX      XX
PF      24-JUL-2001; 2001WO-US023270.
XX      XX
PR      24-JUL-2000; 2000US-00624693.
XX      XX
PA      (PCBU-) PCBU SERVICES INC.
XX      XX
PI      Yoshida RK, Koester AB;
XX      XX
DR      MPI: 2002-268973/31.
XX      XX
DR      P-PSDB; ABB07693.
XX      XX
PT      Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT      treating mammal having disease or disorder from phenylketonuria, cancer,

```

```

PT      human immunodeficiency virus infection and human cytomegalovirus
PT      infection.
XX      XX
XX      Example 3; Fig 2; 135bp; English.
XX      XX
CC      The invention relates to an isolated and purified yeast phenylalanine
CC      ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC      host cell comprising the polynucleotide is useful for the production of L
CC      -PAL, by adding the polynucleotide or the host cell to a composition
CC      comprising trans-cinamic acid, or trans-cinamate and ammonia, and for
CC      improving the production of PAL, its analog or another optically active
CC      unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC      are useful for treating a mammal having a disease, disorder or condition
CC      selected from phenylketonuria, cancer, human immunodeficiency virus
CC      infection and human cytomegalovirus infection. The present sequence
CC      represents a R. toruloides PAL polypeptide encoding cDNA. (Updated on 07-
CC      AUG-2003 to correct OS field.)
XX      XX
SQ      Sequence 2439 BP; 440 A; 920 C; 631 G; 448 T; 0 U; 0 Other;
Query Match      68.7%; Score 1699.4; DB 6; Length 2439;
Best Local Similarity 79.4%; Pred. No. 1,4e-228;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;
QY      1 ATGGCCCCCTCCNTGACTGATCGGACCTCGNTGSCCAACGGCNCNCNACGNTNG 60
Db      1 ATGGACCTCGCTCGACTGATCTCGCATCTGTCGCAACAGCGCTCGCAACG 60
QY      61 CAGCGGNNCCGNNNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCGNNCCG 120
Db      61 CAGGCTGT-----CAATGGCGCTGACCAACTC-----GCAATGCGAG 100
QY      121 GCTCGCTCTCTCCGACCAACCCAGNNAGCGAGCTGACATGTTGAGNAGATCTCTCGCG 180
Db      101 GCTCGACCTCGCCCAACCCAGGTCAGAGGTGACATGTGGAAGAATGCTGCGCG 160
QY      181 ACCCAACCCGNNACGNNACGNNNTGAACTCGAGGCTACACCTTACCTTGGNAGCTCG 240
Db      161 CGCCGACCGACTCGA---CGCTGAACTCGACGCTACTGCTTCACTCGAGACCTCG 217
QY      241 TCGGCGCGCGCNCGCAAGGGGCGCNCNGTCGCGCTCNCAGACAGNCGAGATCGCGCA 300
Db      218 TCTCGGCGCGCAGAGAGGAGGCGGCTGTGCGCTCAAGAGACG--CGACGAGATCGCTCA 276
QY      301 AAGATCGACAAAGNGTGAAGTCTCTCCGNNCCAGCTCNAACAAGNGTCTACGNGTC 360
Db      277 AAGATTGACAAATCGGTGAGATTCTTGCGCTCGCACTCTTCATGACGCTACGCGCTC 336
QY      361 AGAATGTTTCGGGGGCTCGGCGGACCGGACCGGACGTAAGATGMACTGCTCGAAG 420
Db      337 AGGACTGATTTGGGGATCCGACAGACACCGGACGAGAGCGCATCTGCTCAAGAG 396
QY      421 GGNCTCTCGAGCAACAAGCTCTGCGGTGTCGNNCCAGCTCGNCGANTCCTTNGGCTC 480
Db      397 GGTCTCTCGAGCAACAAGCTCTGCGGTGTCGNNCCAGCTCGNCGANTCCTTNGGCTC 456
QY      481 GAGCGGCGCTCGAGAACTCGCTTCCGCTGAGGTGCTCGGCGGCGCATGACCATTCGC 540
Db      457 GGCCTCGGTCTCGAAGACTCGCTTCCCTCGAGGTGTTGCGGCGGCGCATGACCATTCGC 516
QY      541 GTCACCTGCTCAACGCGGCGCACTCGGCGGCTCGGCGGCTCGGCGGCGGCTCGAC 600
Db      517 GTCACCTGCTCAACGCGGCGCACTCGGCGGCTCGGCGGCTCGGCGGCTCGAC 576
QY      601 AACTTCTCAACGAGGATCAACCCCATGTCCTTCGCGGCGGCGGCGGCGGCGGCTCG 660
Db      577 AACTTCTCAACGAGGATCAACCCCATGTCCTTCGCGGCGGCGGCGGCGGCTCG 636
QY      661 GCGGAGCTCTCCGCTCTCTCATATCGCGGCGGCGGCTCGGCGGCGGCGGCGGCGG 720
Db      637 GCGGAGCTCTCTCTCTCTCTCATATGACGCGGCGGCGGCTCGGCGGCGGCGG 696
QY      721 GTCACGCTNNNTCAAGAGGCGGAGGAGGATCATGTGCGCGCGGCGGCGGCGGCTCG 780

```

Db 697 GTGACGTCGTCCACGAGG3CAAGAGAAATCTGTACGCCCGGAG3CATGGCCCTC 756
 Qy 781 TTNGATCTCGAGCCGTCGTCCTCGAGCCGAGAGAG3TCTCGTCTTGTCAACGAGACG 840
 Db 757 TTGAACCTGAGAGCCGTCGTCCTCGAGCCGAGAGAG3TCTCGTCTTGTCAACGAGACG 816
 Qy 841 GCCGTCCTCGAGCCGTCGTCCTCGAGCCGAGAGAG3TCTCGTCTTGTCAACGAGACG 900
 Db 817 GCCGTCCTCGAGCCGTCGTCCTCGAGCCGAGAGAG3TCTCGTCTTGTCAACGAGACG 876
 Qy 901 TCGCAGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCG 960
 Db 877 TCGCAGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCG 936
 Qy 961 CCNTTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 1020
 Db 937 CCCTTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 996
 Qy 1021 CGACGCTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1080
 Db 997 CGACGCTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1056
 Qy 1081 GACGAGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCG 1140
 Db 1057 GACGAGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCG 1116
 Qy 1141 GGCCTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 1200
 Db 1117 GGCCTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 1176
 Qy 1201 TCGACGAGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAG 1260
 Db 1177 TCGACGAGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAG 1236
 Qy 1261 TTCGAGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAG 1320
 Db 1237 TTCGAGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAG 1296
 Qy 1321 GCGACGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAG 1380
 Db 1297 GCGACGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAG 1356
 Qy 1381 CCNTTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 1440
 Db 1357 CCCTTCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 1416
 Qy 1441 GCCGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1500
 Db 1417 GCCGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1476
 Qy 1501 CAGCGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1560
 Db 1477 CAGCGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1536
 Qy 1561 ACNGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1620
 Db 1537 ACNGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1596
 Qy 1621 CAGCGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1680
 Db 1597 CAGCGCGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1656
 Qy 1681 NCNNGCTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCA 1740
 Db 1657 GTCTGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCG 1716
 Qy 1741 CTGAGGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1800
 Db 1717 CTGAGGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1776
 Qy 1801 CTGAGGAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1860

Db 1777 CTGTCGCCGCTGGACGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1836
 Qy 1861 NNCTCTCTGCGNNNNNCCANNAGCTCTGCTCTGCGNCCGCTCAACGCTCTGGAAGTGGCT 1920
 Db 1837 TCGTCGACGTC-----GCTCTGCTCTGCGCGCGCTGCAACGCTGGAAGTGGCGG 1885
 Qy 1921 CCGCGAGAGAGCCATCTGCTCAAGCGCGANAGTCGCGACNCTCTCTGCGNCCGCTCT 1980
 Db 1886 CCGCGAGAGCCATCTGCTCAACGCGCGACGCGCGACGCGCGACGCGCGACGCGCGCTCT 1945
 Qy 1981 CCGCGAGAGCCATCTGCTCAAGCGCGANAGTCGCGACNCTCTCTGCGNCCGCTCT 2040
 Db 1946 CCGCGAGAGCCATCTGCTCAAGCGCGANAGTCGCGACNCTCTCTGCGNCCGCTCT 2005
 Qy 2041 TCGCGAGAGAGCTGGGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
 Db 2006 TCGCGAGAGAGCTGGGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2065
 Qy 2101 TCGCGAGAGAGCTGGGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2160
 Db 2066 TCGCGAGAGAGCTGGGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2125
 Qy 2161 AGCTCTCTGCTCAAGAGTCTGCTTGAAGGCTGCTTGAAGGCTGCTTGAAGGCTGCTTGA 2220
 Db 2126 AGCTCTCTGCTCAAGAGTCTGCTTGAAGGCTGCTTGAAGGCTGCTTGAAGGCTGCTTGA 2185
 Qy 2221 NNCCNN 2280
 Db 2186 TATCCGCGCTGACCTTTAGAGCTGCTTTGTCGAGCTCGAGTCTGCTGCTGCTTCT 2245
 Qy 2281 TNN 2340
 Db 2246 TCGTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2304
 Qy 2341 NGTNN 2400
 Db 2305 AGTACCCTGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2364
 Qy 2401 NNNNNAAA 2408
 Db 2365 CGCATAAA 2372

RESULT 5
 AAD45815
 ID AAD45815 standard; DNA; 2439 BP.
 XX
 AC AAD45815;
 XX
 DT 07-AUG-2003 (revised)
 DT 27-DEC-2002 (first entry)
 XX
 DE Rhodotorula toruloides PAL DNA.
 XX
 KM Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
 KM cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
 KM human cytomegalovirus infection; cytostatic; virucide; gene; ds.
 OS Rhodospiridium toruloides.
 XX
 FH Key Location/Qualifiers
 FT 1.2151
 FT CDS /tag= a
 FT /product= "PAL protein"
 FT /EC_number= "4.3.1.5"
 PN
 XX US20021027.2-A1.
 XX 01-AUG-2002.
 XX 24-AUG-2001; 2001US-00939408.
 XX 24-JUL-2000; 2000US-00624693.
 PR

PR 24-JUL-2001; 2001MO-US023270.

XX (PCBU-) PCBU SERVICES INC.

XX Yoshida RK, Koestera AB;

XX MPI; 2002-690616/74.

DR P-PSDB; AAE27941.

PT Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
XX human immunodeficiency virus or human cytomegalovirus infection.

XX Example 3; Page 41-43; 74pp; English.

CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
CC such proteins. PAL sequences are useful for producing L-phenylalanine.
CC They are useful for treating mammals having diseases, disorders or
CC conditions that would benefit from treatment with PAL proteins such as
CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
CC human cytomegalovirus (HCMV) infection. Sequences of the invention are
CC also used in gene therapy. The present sequence is R. toruloides PAL DNA.
XX (Updated on 07-Aug-2003 to correct OS field.)

SQ Sequence 2439 BP; 440 A; 920 C; 631 G; 448 T; 0 U; 0 Other;

Query Match 68.7%; Score 1699.4; DB 6; Length 2439;
Best Local Similarity 79.4%; Pred. No. 1.4e-228;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

1 ATGGCCCCCTCCNTGACTCGATCGGACCTCGNTCCGCAACGCGCNTCGMAACGGATNG 60
1 ATGGCAACCTCGCTCGATCGATCTCGACCTCGTTCCGMAACGCGCGTCCGATCCGCAAG 60
61 CACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
61 CAGGCTGT-----CAATGCGCGCTCGACCACTTC-----GCAATCGGCGAG 100
121 GGTGCTCTCTCCGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
101 GCTCGACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 160
181 ACCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
161 CGCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 217
241 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
218 TCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
301 AAGATCGACCAANAGGTGAGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
277 AAGATTGACCAATCGGTGAGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 336
361 ACGATGCTGTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
337 ACGATGATTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
421 GCGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
397 GCTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456
481 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
457 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 516
541 GTCATCTCGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
517 GTCACACGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
601 AACTTCTCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660

Db 577 AACTTCTCAACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
QY 661 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
Db 637 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 696
QY 721 GTCACGCTGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 697 GTCACGCTGTTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756
QY 781 TTNAGCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
Db 757 TTCAACTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 816
QY 841 GCGGCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Db 817 GCGGCTCTGAGATGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 876
QY 901 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
Db 877 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 936
QY 961 CCNTTCTCAACGAGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Db 937 CCNTTCTCAACGAGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 996
QY 1021 CGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
Db 997 CGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1056
QY 1081 GACGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Db 1057 GACGACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1116
QY 1141 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
Db 1117 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
QY 1201 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
Db 1177 TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1236
QY 1261 TTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
Db 1237 TTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1296
QY 1321 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
Db 1297 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1356
QY 1381 CCNTTCTCAACGAGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
Db 1357 CCNTTCTCAACGAGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1416
QY 1441 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
Db 1417 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
QY 1501 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
Db 1477 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1536
QY 1561 ACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
Db 1537 ACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1596
QY 1621 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
Db 1597 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1656
QY 1681 CCNTTCTCAACGAGTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740
Db 1657 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716

QY 1741 CTGNGGACAGAGTCAACAAGCGCTCAACAAGCGCTCGAGCAGACCAACTGCTAGAC 1800
 DB 1717 CTGTCGAGAGAGTAAACAAGACGCTCCCAAGCGCTCGAGCAGACCAACTGCTAGAC 1776
 QY 1801 CTGAGCGCGCTGGAGAGAGCTTCTGTTCTGCGAGACGGACCGCTGCTGAGANNNNN 1860
 DB 1777 CTGTCGCGCGCTGGAGAGAGCTTCTGTTCTGCGAGACGGACCGCTGCTGAGAGTCTC 1836
 QY 1861 NNGTCTCGCCNNNNNCCANNAGTCTGCTGCGAGCGCTGACCGCTGAGAGTGGCT 1920
 DB 1837 TCGTCGACGCTC-----GCTCGCTCGCGCGCTGACAGCTGAGAGGTGCGCG 1885
 QY 1921 CGCGCGAGAGGCGCATCTGCTCAGCGCGANNNGTCGCGAGACGCTTCTGAGCGCGCGT 1980
 DB 1886 CGCGCGAGAGGCGCATCTGCTCAGCGCGANNNGTCGCGAGACGCTTCTGAGCGCGCGT 1945
 QY 1981 CGTCTGCTGCGCGCGCTCAGCTCTGCGCGCGAGCGCGCTGCTGCTGCTGCTGCTG 2040
 DB 1946 CGAGCTGCTGCGCGCGCTCAGCTCTGCGCGCGAGCGCGCTGCTGCTGCTGCTGCTG 2005
 QY 2041 TCGCGAGAGAGCTGCGCGCTCAGAGCGCGCGCGCGAGAGTCTTCTGCGAGAGAGG 2100
 DB 2006 TCGCGAGAGAGCTGCGCGCTCAGAGCGCGCGCGCGAGAGTCTTCTGCGAGAGAGG 2065
 QY 2101 TGACGATCGGACCAACGCTCTCCGATCTAGAGGCGCATCAAGNCGCGCATCAACC 2160
 DB 2066 TGACGATCGGCTCGAAGCTCTCAAGATCTAGAGGCGCATCAAGTCCGCGAGATCAACA 2125
 QY 2161 AGCTCTGCTCAAGATCTCGCTAGAGNCCNNNNNCCNNNNCTGCTGCTGCTGCTGCTG 2220
 DB 2126 AGCTCTGCTCAAGATCTCGCTAGAGATCTTCCGCTGCTGCTGCTGCTGCTGCTGCTG 2185
 QY 2221 NNNCCNNNNNNCTTGTGNNNNCTGNNNNCTGNNNNNNNNNNNNNNNNNNNNNN 2280
 DB 2186 TATCCGCGCTGATCTTAGAGACTGCTTCTGTCGAGCTGAGATCTGCGATCTGCTTCT 2245
 QY 2281 TNNNNNNNNNNCTTNNNNCTCNCNNNNNNNNNNCTNNNNNNNNNNNNNNNNNNNN 2340
 DB 2246 TCGTCTTGGCTGCTGCTC-TAGACCGTGTCCGATTAACCTCGAGATGTAATACAAAG 2304
 QY 2341 NCTNNNNNNNNNNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2400
 DB 2305 AGTACCATTCACGAGATCGATTAATCAAGAGAGATCTAGGCTTGCGGAGGCTTCTG 2364
 QY 2401 NNNNAAA 2408
 DB 2365 CGCATAAA 2372
 RESULT 6
 ID AAN81101 standard; DNA; 2331 BP.
 NC AAN81101.
 XX 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 19-NOV-1990 (first entry)
 DE Sequence encoding L-phenylalanine ammonia lyase (PAL) from Rhodospiridium
 turoroides.
 XX Cinnamic acid; expression vector; MT 10414 Ferump-8976; pYtp6; se.
 OS Rhodospiridium tororoides.
 XX BP278706-A.
 PD 17-AUG-1988.
 PF 06-FEB-1988; 88BP-00301011.
 XX

PR 06-FEB-1987; 87JP-00024705.
 PR 18-JUN-1987; 87JP-00152357.
 PA (MITK) MITSUI TOANSU CHEM INC.
 XX Fukuhara N, Yoshino S, Sone S, Nakajima Y, Makiguchi N;
 DR WPI, 1986-229543/33.
 XX
 PT Recombinant plasmid for expression of L-phenylalanine ammonia-lyase -
 PT having combined promoter of tac promoter and P1 promoter for enhanced
 PT expression.
 XX
 PS Example; Fig 2 (A)-2(D); 37pp; English.
 CC Plasmids were from E.coli MT 10414 (Ferump-8976) carrying the plasmid
 CC pYtp6 contg. the structural gene for PAL from Rhodospiridium tororoides.
 CC Inventors claim a recombinant plasmid contg. a DNA sequence coding for
 CC PAL with the SQ in AAP0501. Procedure for cloning the structural gene
 CC for PAL (AAN81101) is described as a reference example. The hybrid
 CC plasmid permits more efficient expression of PAL in E.coli. PAL is used
 CC for the prodn. of L-phenylalanine from cinnamic acid and ammonia (Updated
 CC on 25-MAR-2003 to correct PR field.) (Updated on 27-AUG-2003 to correct
 CC OS field.)
 CC
 SQ Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;
 Query Match 67.8% Score 1677.4; DB 1; Length 2331;
 Best Local Similarity 80.4% Pred. No. 1.6e-225;
 Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;
 QY 1 ATGGCCCCCTGCNTGACTGATCGAGCACTGCTGCTGCGCAAGCGCTCNCNAAAGGNTNG 60
 DB 1 ATGGACCCCTGCTGCACTGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 60
 QY 61 CAGCGTGT-----CATGGCGCTTGCACCACTC-----GAGTGCAG 100
 DB 61 CAGCGTGT-----CATGGCGCTTGCACCACTC-----GAGTGCAG 100
 QY 121 GCTGCTCTCTCCGACACCGACGAGNNGACGAGCTGCACTGCTGAGAGNAGATCTGCGCG 180
 DB 101 GCTGCACTCTCCGACCAACCGAGTCAAGAGTCAAGATCTGCAAGAGATCTGCGCG 160
 QY 181 ACCCAACGANNAGGANNNTGAACTCGAGCGGATACCTTCACTCTGCGAGAGTGTG 240
 DB 161 CGCCGACCGACTGCA---CGCTGCACTGCAAGGCTTCACTGCACTCTGCGAGAGTGTG 217
 QY 241 TCGGCGCGCGCGCAAGGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 DB 218 TCTGCGCGCGCGAGAGGAGGAGGCTGTCCGCTGCTGCAAGGAG-CGACGAGATCGCTCA 276
 QY 301 AAGATCGACAAAGNGTCAAGTCTCTCGANNCCAGCTCAACAACAGNGTCTACGAGTGC 360
 DB 277 AAGATTGACAAATGCTGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 336
 QY 361 ACGACTGCTTGGCGGCTGCGGCGCGACCGAGCTGAGATGATGATGATGATGATGATGATG 420
 DB 337 ACGACTGATTTGGCGGATTCGCGAGACACCGACGAGAGACCATCTGCTGCTGCGAGAG 396
 QY 421 GCTCTCTCGAGACCAAGCTCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
 DB 397 GCTCTCTCGAGACCAAGCTCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 456
 QY 481 GCGCGCGCGCTCGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
 DB 457 GCGCGCGCTCGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 516
 QY 541 GTCAACTGCTCAACGCGGCACTCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
 DB 517 GTCAACAGCTTGAACCGGCGCACTCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 576
 QY 601 AACTTCTCAACCAAGGATACCCCATATGTCCTTCCGCGGCAACATCTTGGGCTGTG 660

Db	577	AACTTCCTCAACCAAGGATACCCCATGTCCTCCCTCGGAGGACCAATCTGTCGCG	636
QY	661	GGCGACTCTCCCNCTCTNTAATGCGCGGCCCATACCGGTACCCGGAAMNAA	720
Db	637	GGCGACTCTCTCTCTCTCTCTCAATGACGCGCATACGAGGTACCCGGAACGA	696
QY	721	GNCAAGTNNMCAAGAGGAGANNAGAAATATGTCGCCCGGAGGCAATGCGCGTC	780
Db	697	GTGCAGTGTCTCAAGAGGAGCAAGAAATCTGTACGCCCGGAGGCAATGAGCGCTC	756
QY	781	TTNGTCTCGAACCCTGTGTCCTTGCGGCCGAAGAGGTCTCGTCTCGTCAACGGCAG	840
Db	757	TTCAACCTCGAACCCTGTGTCCTTGCGGCCGAAGAGGTCTCGTCTCGTCAACGGCAG	816
QY	841	GCCGTCCTCGCTCGATGGCAACCTTGCTGTGACAGACGACATGTCTCGCTCTC	900
Db	817	GCGCTTACGATCATATGGCCACCTCGCTGTGACAGACGACATGTCTCGCTCTC	876
QY	901	TGCAAGGCGCTCACGCGCTCTACCGCTGAGAGCATGGTCGCGCACCGCGCTCTCAC	960
Db	877	TGCAAGTGTCTCAGGCGCATATCGGTGAAAGCATGTGTGGCCACGCGGCTCGTCCAC	936
QY	961	CNNTTCTTCAAGAGCTACGCGCTCTACCCGACCCATGATGAGGTGCGCGCAATCT	1020
Db	937	CCCTTCTTCAAGAGCTACGCGCTCTACCCGACCCATGATGAGGTGCGCGCAATCT	996
QY	1021	CGACGCTCTCGAGGAGCGAGNNGTGTCGCTCCACACAGAGAGAGGTCAAGTCAAG	1080
Db	997	CGCAAGCTCTCGAGGAGCGAGNNGTGTCGCTCCACACAGAGAGAGGTCAAGTCAAG	1056
QY	1081	GACGACGAGGCGATTTCTCCGCCAGACCGCTACCCGCTCGACGTCGCTCAATGCTC	1140
Db	1057	GACGACGAGGCGATTTCTCCGCCAGACCGCTACCCCTTGACAGCTCTCTCAAGTGGTC	1116
QY	1141	GCGCGCTGTGAGGACATGATTCAGGCCACGCGCTCTCTCGCTCGAGGCGCGAG	1200
Db	1117	GCGCGCTGTGAGGACATTCATTCAGGCCACGCGCTCTCTCAACATGAGGCGCGAG	1176
QY	1201	TGACGACCGCAACCCGCTCATGACGTGAGAACAGAGAACCCACCGCGCGCAAC	1260
Db	1177	TGACGACCGCAACCCCTCTCATGACGTGAGAGAACAGCTTCGACACCGCGCGCAAT	1236
QY	1261	TTCCAGGCGCGCTGTGCGNAACAGCATGAGAGATCTGCGTCCGCTCGGCTGTGATC	1320
Db	1237	TTCCAGGCTGCGCGTGTGACCAACAGTGAAGATCTGCTCGGCTGTGCGCTGTGATC	1296
QY	1321	GGCAAGCTTAATTACGACAGCTACCGAGATGTCAACGCGGCAATGAACGCGGCTTN	1380
Db	1297	GGCAAGCTTAATTACGACAGCTACCGAGATGTCAACGCGGCAATGAACGCGGCTTC	1356
QY	1381	CNCTCTGCTGTGCTGCGGAGAACCTCGTCTCTATATCTGCAAGGCGCTGACATT	1440
Db	1357	CCCTCTGCTGTGCGCGCGAGAACCCCTCGTCTCTCAACATGCAAGGCGCTTCAACATC	1416
QY	1441	GCGCGNGNGNTACATTCGGAAGTGGANACTTCCAAACCGGTYNACATCAACCTC	1500
Db	1417	GCGCGTGGCGGTAACTCTGGAAGTGGAGACCTGCGCAACCTGTGACGACGATGTC	1476
QY	1501	CAGCGGAGAGATGAGCAACAGGCGGTCAACTCGCTCGNCTCATCTCGAGNCGCGC	1560
Db	1477	CAGCGGCTGAGATGAGCAACAGGCGGTCAACTCGCTCGNCTCATCTCGGCTGTGCGC	1536
QY	1561	ACNGCGGAGGCAAGAGTCTTCTCTCTCTGCGCACCACTGACTACTGCGGTC	1620
Db	1537	ACGACCGAGTCAAGAGTCTTCTCTCTCTGCGCACCACTGACTACTGCGGTC	1596
QY	1621	CAGCGGTGACCTCGCGCGATGAGTTCGAGTTCAAGAGAGTTGACCCGANTNTC	1680
Db	1597	CAGCGCATGACTTGCGCGCGATGAGTTCGAGTTCAAGAGAGTTTGCGCCGACGATCTC	1656
QY	1681	NCNNGCTCTNACAGACGACTTTGGCNCNGCCTCGACGCGNNNNNNNNNNNACAA	1740
Db	1657	GTCTGCTATGACACGACTTTGTCTCTCGCATGACGCGCTCGAACCTGCGGACGAG	1716

QY 1744 CTCGAGGACAAGGTCTCAACAAGNGCCTCMAACAACGCTGAGCAGACCAACTCGTAGCAC 1804
 Db 1717 CTCGTCTGGAAGGTGAATAAGACGCTTCGCAAGCGCTTGAGAGACAACTCTGTACAC 1776
 QY 1801 CTCGAGCCGCGGTGGACAAGACGCTTCTGTCTTGCCACACGGCACCCTGTCGAGNNNNN 1860
 Db 1777 CTCGTCTCCGCGCTGGACAGACGCTTCTCTTTCGCGCGGACCGTCTGTAGAGTCTCTC 1836
 QY 1861 NNGTCTCTGCCNNNNGCCANNAGTCTCTGCTCGCAGCCCTCAAACGCTTGAAGGTCTCT 1920
 Db 1837 TCGTCGACGTC-----GCTCTGCTCCCGCGCTCAACGCTGGAAGGTCTCGC 1885
 QY 1921 CCGCGGAGAAGGCCATCTCGCTCACGGGCAAGTCCGCGACACGCTTCTGAGCGCCGCT 1980
 Db 1886 CCGCGGAGTCGCCCATCTCGCTCACCGCCAACTCCGCAAGCTTCTGTGTCGCGGCT 1945
 QY 1981 CGTCGTCTGCGCCGCGCTCNCGTACTCTCGCGCGCAAGCGCGCTGTACTGTCTG 2040
 Db 1946 CGACCTCTGTCGCGCGGCTCTGTACTCTTCGCGCGCACTAGATCTGTACGCTTCTG 2005
 QY 2041 TCCGCGAGGAGCTCGGCGCTCAAGCGCCGCGCGGCAAGTCTTCTCGGCAAGCAGAGG 2100
 Db 2006 TCCGCGAGGAGCTTGGCTCAAGGCTCCGCGCGCAAGAGCTCTTCTCGCAAGAGAGG 2065
 QY 2101 TCACGATCGGACCAAGAGCTCCCGCATCTACGAGGCCATCAAGNNCGCNCATCAAC 2160
 Db 2066 TCACGATCGGCTCGACAGCTCTCCAAAGTATCAAGAGGCCATCAAGTCGGGCAAGATCA 2125
 QY 2161 AGCTCTCTGTCAGAATGCTCGCTAGNNNCNNNNNCNNNNNCNNNNNCNNNNCNC 2220
 Db 2126 AGCTCTCTCTCAAGATGCTCGCTTAGACACTCTTCCCACTCTCGCATCCCTTCATACC 2185
 QY 2221 NNCCNNNNNNNNCTTTTNGNNNTCGNNTCGNNTCGNNNNCGGANNNTNNCNNNNNNNN 2280
 Db 2186 TATCCGCGCTGACACTTAGAGCTCGCTCTTGTGAGACTCGGATCTCGCATCGCTTCT 2245
 QY 2281 TNNNNCTNNCTTNNCTCNCNNNNANCNNGCTNNNTNNNNCTNNNGTNNNNNNNNNNC 2340
 Db 2246 TCGTCTTGTGCGCTCTCTC-TAGACCGTGTGGATTACTCGAAGTGTGAATCAAGC 2304
 QY 2341 NGTNNNCANNNA 2352
 Db 2305 AGTACCCATCCA 2316
 RESULT 7
 AAN81116
 ID AAN81116 standard; DNA; 2331 BP.
 AC AAN81116;
 XX 24-OCT-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 12-NOV-1990 (first entry)
 XX L-phenylalanine ammonialyase structural gene.
 DE L-phenylalanine ammonialyase; PAl; foreign gene expression;
 KM L-phenylalanine ammonialyase; PAl; foreign gene expression;
 KM culture temperature; expression regulation; ss.
 XX Rhodospiridium rubroideum; (IFO 559) .
 OS
 FH Key Location/Qualifiers
 FT CDS 1. 2151
 FT /tag= a
 FT /Product= "PAl"
 PN EP279665-A.
 XX 24-AUG-1988.
 XX 18-FEB-1988 88EP-00301356.

XX 19-FEB-1987; 87JP-00034397.
PR 18-JUN-1987; 87JP-00152359.
XX
PA (MITK) MITSUI TOATSU CHEM INC.
PI Fukushima N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y;
XX WPI; 1988-236895/34.
DR P-PSDB; AAP80513.
XX
PT Regulation of expression of foreign gene in Escherichia coli - by
PS maintaining temp. of culture at 40 deg. C or more to suppress expression.
XX
PS Disclosure;); 36pp; English.
CC The cloned DNA from pSM13 isolated from the cloned cell carrying pSM13
CC was digested with various restriction endonucleases. The sequences from
CC the fragments were analysed by the method of Maxam-Gilbert and also by
CC biochemical means using the method of Maat. The sequences obtained were
CC edited using the GENAS program produced by Mitsui Information Development
CC Co. The sequence so determined is represented here. The regulation of
CC expression of this foreign gene in E.coli carrying a recombinant plasmid
CC comprises maintaining the culture at at least 40 deg. C. See also EP-
CC 279664. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
CC 2003 to correct PI field.) (Updated on 24-OCT-2003 to standardise OS
CC field)
XX
SQ Sequence 2331 BP; 426 A; 888 C; 592 G; 425 T; 0 U; 0 Other;
Query Match 67.8%; Score 1677.4; DB 1; Length 2331;
Best Local Similarity 80.4%; Pred. No. 1.6e-225;
Matches 1891; Conservative 0; Mismatches 425; Indels 36; Gaps 6;
QY 1 ATGGCCCCCTCCCTGACCTGATTCGGACCTGATTCGGACCAAGGCTTCCGAGGTTG 60
DB 1 ATGGCACTCTGCTGACCTGATTCCTGACCTGTTTGGCAAAAGGCTGCAATCCGCAAG 60
QY 61 CAGCCGNNCCGNNNGNCGNACGCGGCGCCACGCTCAAGCTCNGNCGCCGCGNG 120
DB 61 CAGGCTGT-----CATGCGCTCTGACCAACCTC-----GCACTTCGAG 100
QY 121 GCTGCTCTCTCCGACCACTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
DB 101 GCTGCACTGCTCCGACCAACCTCAGGTCAGGAGGAGGAGGAGGAGGAGGAGGAGG 160
QY 181 ACCCGACGNNACGNNACGNNNTGAACTGAGCGGAGTACACCTCACTCGAGAGCTG 240
DB 161 CGCGCAACGACTGA---CGCTGAACTGAGGCTTACTGCTCAACTCGAGAGAGCTG 217
QY 241 TCGGCGCCGCGCGCAAGGCGCGCGCGCGCTCCGCTCAGAGAGAGAGAGAGAGAG 300
DB 218 TCTCGCGCCCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 276
QY 301 AAGATGAGCAANAAGTGAAGTTCTCCGNNCNCAGCTCNAACAAGAGTCTAGCGGTC 360
DB 277 AAGATTGACAAATCGGTGAGATTCTTGAGCTGCAACTCTTCAAGAGGCTTACGAGGCTC 336
QY 361 ACGACTGATTTGGCGGCTCGGCGCAACCCGGAATGAGATGCAATCTCGCTCAGAG 420
DB 337 ACGACTGATTTGGCGGATTCGCGAGACACCGCGCACCGAGACGCAATCTCGCTCAGAG 396
QY 421 GCGTCTCTGAGAGCACTGCTCTGCGGTGTCCTCCNAGTGCAGTCTTCCGCTC 480
DB 397 GCTCTCTGAGAGCACTGCTCTGCGGTGTCCTCCGCTTCTGCTTCTGCTTCCGCTC 456
QY 481 GAGCGGCGCTCGAGAACTCGCTTCGCTCGAGGTCGTCGCGGCGCGCATGACCATCGG 540
DB 457 GCGCGGCTCTGAGAACTCGCTTCCCTCGAGGTGTTTCCGCGGCGCATGACCATCGG 516
QY 541 GTCAACTGCTCAACGCGGCGCACTCGGCGTTCGCTCTGCTCTGAGCGCTGACC 600
DB 517 GTCAACAGCTTGAACCGCGGCGCACTCGGCTGTCGCTCTGCTCTGAGGCGCTGACC 576

QY 601 AACTTCCTCAACGAGGATGACCCCCATGCTGCCCTCCGCGGACCATCTCGGCGCTG 660
DB 577 AACTTCCTCAACGAGGATGACCCCCATGCTGCCCTCCGCGGACCATCTCGGCGCTG 636
QY 661 GCGGACTCTCCGCTCTCTCTATGAGTCCGCGCGGACATCAACGAGTCAACCGGACGNNCAG 720
DB 637 GCGGACT 696
QY 721 GTNCACTGNNTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
DB 697 GTGCACTGCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
QY 781 TTNAGTCTGAGGCGGCTGCTCTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
DB 757 TTCAACTCTGAGCGGCTGCTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 816
QY 841 GCGTCTCCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
DB 817 GCGTCTGAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 876
QY 901 TCGGAGGCGCTCAAGGCTCTMAAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
DB 877 TCGAGTGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 936
QY 961 CCNTTCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
DB 937 CCCTTCTTCAAGAGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996
QY 1021 CGGAGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 997 CGGAGCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
QY 1081 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
DB 1057 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1116
QY 1141 GCGCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
DB 1117 GCGCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
QY 1201 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
DB 1177 TCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
QY 1261 TTCCAGGCGGCGCTGTCGNNACGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
DB 1237 TTCCAGGCTGCGGCTGTCGNNACGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1296
QY 1321 GCGAGGCTCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
DB 1297 GCGAGGCTCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1356
QY 1381 CCNTTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
DB 1357 CCCTTCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
QY 1441 GCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1417 GCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1476
QY 1501 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
DB 1477 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536
QY 1561 ACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
DB 1537 ACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1596
QY 1621 CAGGCGCTGAGCTTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
DB 1597 CAGGCGCTGAGCTTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1656

```
QY 1681 NCNNGCTGCTNTNAGACGACTTTGGCNCNGCCCTCGACGCGCNNNNNNNNNNACGAA 1740
Db 1657 GTCCTCGTATGACGACGACTTTGGCTCCGCGATGACCGGCTCGAACTCGCGAGAG 1716
QY 1741 CTCGNGGACGAAGGTACAAGCGCTCNAACAAGCGCTCGAGAGACCACTCGTAAGAC 1800
Db 1717 CTCGTGAGAGAGGTGACAGACGCTCGCCAGCGCCTCGAGACCACTCGTAAGAC 1776
QY 1801 CTCGAGCCCGCGCTGGACGACGCGCTTCGTCTCGGACCGGACCGCTCGTGAAGTCCGCT 1860
Db 1777 CTCGTCCCGCGCTGGACGACGCGCTTCGTCTCGGACCGGACCGCTCGTGAAGTCCGCT 1836
QY 1861 NNCTCTCGCCNNNNGCCANNAGTCTCGCTCGCNGCCGCTCAAGCGCTGGAAGTCCGCT 1920
Db 1837 TCCTGACGCTC-----GCTCTGCTCGCGCGCTCAAGCGCTGGAAGTCCGCT 1885
QY 1921 CGCCGACGAAGGACATCTCTGCTACGCGCNGANGTCCGCGACNCTCTGAGCGCNCGCT 1980
Db 1886 CGCCGACGAGTGGGACATCTCTGCTACCGCGCAAGTCCGCGACACCTCTGCTCCCGCGCT 1945
QY 1981 CGTGTGTCGCGCCGCGCTCNCGTACTCTCGCGCGGACGCGCGCTCTGTAATCTTTCG 2040
Db 1946 CGACCTGTGCGCCGCGCTCTGTACTCTCTCGCGCGGACGCTGATCTCTACGCGCTTCG 2005
QY 2041 TCAGGAGAGAGCTCGGCGCTCAAGGCGCGCGCGGAGAGCTCTCTCGGCAAGGAGAGG 2100
Db 2006 TCAGGAGAGAGCTCGGCGCTCAAGGCGCGCGCGGAGAGAGCTCTCTCGGCAAGGAGAGG 2065
QY 2101 TGACGATCGGACCAAGCTCTCTCGCATCTACAGGCGCATCAAGNCGCGCAGATCAACC 2160
Db 2066 TGACGATCGGCTGGAAGCTCTCTCAAGATCTACAGGCGCATCAAGTCCGCGGAGATCAACA 2125
QY 2161 AGCTCTCTGTAAGATGCTCGCTGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2220
Db 2126 AGCTCTCTCAAGATGCTCGCTGTAAGACACTCTCTCCACTCTCGGATCCCTTCATACCC 2185
QY 2221 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2280
Db 2186 TATCCGCGCTGACACTTTAGAGACTGCTCTTGTGAGACTGGAATCGCATCCCTCTT 2245
QY 2281 TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
Db 2246 TCGTCTCTGCTGCGCTCTC-TAGACCGTGTGCTATTACTTGAGATTGTGATATCAAGC 2304
QY 2341 NGTNNNNNNNN 2352
Db 2305 AGTACCATTCCA 2316

RESULT 8
ID AAN81399 standard; cDNA; 2328 BP.
XX AAN81399;
XX AC
XX 12-DEC-1990 (first entry)
XX DE Sequence of Rhodospiridium toruloides IFO 559 cells cDNA encoding L-
XX DE Phenylalanine ammonia lyase (PAL).
XX KM Aspartame; artificial sweetener; enzyme; ss.
XX OS Rhodospiridium toruloides.
XX FT Key Location/Qualifiers
XX FT CDS 1..2151
XX FT misc_feature 1..2148
XX FT /tag= a
XX FT /tag= b
XX FT /note= "Claimed"
XX PN BP260919-A.
```

```
PD 23-MAR-1988.
XX 15-SEP-1987; 87EP-00308119.
XX 16-SEP-1986; 86JP-00215864.
XX (MITR ) MITSUI TOATSU CHEM INC.
XX Fukuhara N, Yoshino S, Yamamoto K, Se T, Sone S, Nakajima Y,
XX PI Suzuki M, Nakiguchi N;
XX MPI: 1988-079223/12.
XX DR P-PSDB; AAP81099.
XX PT New L-phenylalanine ammonia lyase and encoding DNA sequence - useful for
XX PT producing L-phenylalanine from cinamic acid and ammonia.
XX PS Disclosure; 9; 37bp; English.
XX CC PAL is useful in prodn. of L-phenylalanine from cinamic acid and an NH3
XX CC donor (process claimed). L-Phe is an essential AA and starting material
XX CC for sweetener aspartame. E.coli transformants will produce PAL without
XX CC induction by expensive AAs; no surfactant is needed to increase cell wall
XX CC permeability.
XX SQ Sequence 2328 BP; 423 A; 889 C; 591 G; 425 T; 0 U; 0 Other;

Query Match 67.7%; Score 1675.8; DB 1; Length 2328;
Best Local Similarity 80.4%; Pred. No. 2.7e-225;
Matches 1890; Conservative 0; Mismatches 426; Indels 36; Gaps 6;

QY 1 ATGGGCCCTCCCTCCTGACATCGATCGGCGACCTCGMTGCGCAAGCGCTCNCMAACGAGTNG 60
Db 1 ATGGGACCCCTCGCTGACATCGATCTCGGACTGTGTGCAAGCGGCTGCTCGCAAG 60
QY 61 CAGGCGGNNCCGNCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
Db 61 CAGGCGTGT-----CAATGGCGCTCGACCAACTTC-----GCACTGCCAG 100
QY 121 GCTGCTCTCTCCCGACCAACGANNAGAGAGCTGAGATGTGAGATGATCTCTGCGCG 180
Db 101 GCTGCGACTCTGCCCAACCAACGAGTACGAGGTGATGTGAGATGTGAGATGTGCGCG 160
QY 181 ACCCCACGNNACGACGANNNTGAACTCGACGCGGTACACCTCGACCTCGGAGCGTGC 240
Db 161 CGCCGACGACTCGA---CGCTGAACTCGAAGGCTACTGCTCACTCGGAGACGTCG 217
QY 241 TCGGCGCGCGCNGCAGAAGGCGCNCNCTCGCGCTGTCAGACAGACGAGATCCGCGCA 300
Db 218 TCTCGGCGCGCGAGAGGCGAGGCGCTGTCCGCTGTAAGAGAG- CGAGAGATCCGCTCA 276
QY 301 AAGATCGACCAANAGGTGAGTCTCTCGANNCCAGCTCNAACAGAGTCTACGAGTGC 360
Db 277 AAGATTGACAAATCGGTGAGTCTCTGCTGCTCGCACTCTCATGAGCGTCTACGGGCTC 336
QY 361 ACGACTGTTTCGGCGGCTCGGCGCACACCGGACTGAGATGAGATCTGCTCCAGAG 420
Db 337 ACGACTGATTTCGGGATTCGCGACAGACACCGGACCGGAGCGCATCTCGCTCAGAG 396
QY 421 GGNCTCTGAGACACCACTCTGCGGTGTTNCTCCNAGTGGANTGANTCTTCCGCTC 480
Db 397 GCTCTCTGAGACCACTCTGCGGTGTTTCTCTCTGTTGTTGATCTCTTCCGCTC 456
QY 481 GGNCGGCGCTCGAGAACTCGCTTCGCTGAGTGTGCGCGGCGCATATGATCGATCGC 540
Db 457 GGGCGGCGCTCGAGAACTCGCTTCCTCGAGGTGTGTGCGGCGCATATGATCGATCGC 516
QY 541 GTCAACTGCTCAGCNGGCGCACTCGGCGTCTGCTCTGCTCTGAGGCGCTCAAC 600
Db 517 GTCAACAGCTGACCCGCGGCGCACTCGGCTGTGCGCTCTGCTCTGAGGCGCTCAAC 576
QY 601 AACTTCTCAACCAAGGATACCCCGATCGTCCCGCTCGCGGACCATCTCGGCGTGC 660
```


XX 24-JUL-2000; 2000US-00624693.
XX (PCBU-) PCBU SERVICES INC.
XX Yoshida RK, Koostra AB;
XX WPI: 2002-268973/31.
XX P-PSDB; ABB07690.
XX
XX Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT treating mammal having disease or disorder from phenylketonuria, cancer,
PT human immunodeficiency virus infection and human cytomegalovirus
PT infection.
XX
XX Claim 4; Fig 2; 135pp; English.
XX
XX The invention relates to an isolated and purified yeast phenylalanine
CC ammonia lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC host cell comprising the polynucleotide is useful for the production of L
CC -PAL, by adding the polynucleotide or the host cell to a composition
CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC improving the production of PAL, its analog or another optically active
CC unnatural amino acid having a PAL-like structure. The PAL polynucleotides
CC are useful for treating a mammal having a disease, disorder or condition
CC selected from phenylketonuria, cancer, human immunodeficiency virus
CC infection and human cytomegalovirus infection. The present sequence
CC represents a R. graninis PAL polypeptide encoding cDNA
XX
XX Sequence 2419 BP; 416 A; 909 C; 659 G; 434 T; 0 U; 1 Other;
SQ
Query Match 65.8%; Score 1629; DB 6; Length 2419;
Best Local Similarity 78.0%; Pred. No. 9.1e-219;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;
QY 1 ATGGCCCTCCCTCCNTGACTGATCGGACCTGNTGCCAAGCGGCTNCNAAAGGNTG 60
DB 37 ATGGCCCTCCCTCCNTGACTGCTCGCCACACAGCTCGCCAAAGGCTTACCAAGGCTCG 96
QY 61 CAGGCCGNNCGNCCNNGNCGNCAAGGGCGGCGACGTCACNCTGNGGCGCGNCGNG 120
DB 97 CAGCGGCTCCGACCAAGTCGGCTGCGGGCCCACTTGGCTCTCCGCGCAAGC--CG 154
QY 121 GCTCGCTCTCTCCGACCCAGNNNGACGACGCTGACATCGTNGAGNAGATCCTCGCG 180
DB 155 GCTTCGATGACGACGCGCGGACAGTGCAGCTCGATCGTGAAGAGAGCTCCTCAGCG 214
QY 181 ACCCGACCGNACGNNTCGAACTCGAGGGTACACCTCACTCCCTCGNAGACGTG 240
DB 215 ACCCGACCGAC--GACGTGCTGAGCTAGCGGGTACAGCTCACTCGTACGTG 271
QY 241 TCGGCGCGCGNCGCAAGGCGCGNCGTCGCGTCNCAAGNCGAGAGATCCGCGCA 300
DB 272 TCGGCGCGCGCGCAAGGGCGGAGGTCGCGTC--CAGAAAGAGAGATCCGCGCA 330
QY 301 AAGATGACAAANAGTCTGAGTTCTCCGNNCGAGCTCNACAGAGTCTAAGAGTC 360
DB 331 CGGTGGAAGAAGCGCTGCACTTCTCAAGGCCAGCTTCAGAACTCGTGTAGAGGTC 390
QY 361 ACAGCTGATTGCGCGCTCGGCGGACACCGGAGTGAAGATGATCTCGTCCAGAG 420
DB 391 ACCACGGGTTTGGTGGCTCGGCGGACAGAGAGATGAAGATGAGCTCAGAG 450
QY 421 GNCCTCTCGAGACAGAGCTCTGCGGTGTCTCCNACGTCGTCGANTCCTTCNGCTC 480
DB 451 GCGCTCATGAGACAGAGCTCTGCGGCTGACGCGGACGTCGTCGCTCTTAGAGCTC 510
QY 481 GAGCGGGGCTCGAGAGCTCGCTTCGCGTGAAGTGTCCGCGGCGCATACATCGCG 540
DB 511 GAGCGGGGCTCGAGAGCTCGCTTCGCGTGAAGTGTCCGCGGCGCATGATGATCGC 570
QY 541 GTCACCTGCTACNCGCGGCACTCGGCGTCCGCTGTCGTCCTCGAGAGCTCAGC 600

DB 571 GTCACCTGCTACNCGCGGCACTCGGCGCTCCGCTCGTCGTCCTTAGAGCGCTCAGC 630
QY 601 AACTTCTCAACACGAGCATACCCCATGTCCTCCCTCGCGGACCATATTCGGGCTCG 660
DB 631 AACTTCTGAACACGAGCATACGCGCATGTCCTCCCTCGCGGCTCATATTCGGGCTCG 690
QY 661 GAGGACCTCTCCGCTCTGTCATGAGCGCGCGGCGCATACCGGTCACCGGAGNNCAAG 720
DB 691 GCGGACCTCAGCGCGCTCTGTCATGAGCGCGGCGGCGCATACCGGTCACCGGAGTCAAG 750
QY 721 GTNCACTGNTNCAAGAGGAGGAGGATCATGTNCGCGCGGAGGCGATCGCGCTC 780
DB 751 GTTCACTGTTTGAAGAGGAGAACGAGAGATCATGTTTGCGCGGAGGCGATCTCCCTC 810
QY 781 TTNAGTCTGAGCGCGCTGCTCGCTCGCGCGGAGGAGGCTCTGTCACAGGAGC 840
DB 811 TTTGATCTGAGGAGCGATGCTCTCGCGCGGAGGAGGCTCTGTCACAGGAGC 870
QY 841 GCGGCTCGCGCTCGATGAGGAGCGCTCTGTCAGAGGAGCATGCTCTCGCTCTC 900
DB 871 GCGGCTCGCGCTCGATGAGGAGCGCTCTGTCAGAGGAGCATGCTCTCGCTCTC 930
QY 901 TCGCAGGCGCTCAGCGCTCTTACAGGTGAGGCGCATGTCGCGCACCGCGCTCGTCCAC 960
DB 931 TCGCAGGCGCTTGAAGCGCTCTCAGCGGTGAGGCGCATGTCGCGCACAGGCTCGTTCGCG 990
QY 961 CANTTCTTCAAGAGTACAGCGCGCTCAGCGCGGAGGAGGATGAGTTCGCGGAGCATC 1020
DB 991 CCGTTCATCAGAGAGTTCGCGCGCGGAGGAGGAGGATGAGTTCGCGGAGCATC 1050
QY 1021 GCGACGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 1051 CCGACGCTCTTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1110
QY 1081 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
DB 1111 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1170
QY 1141 GCGCGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
DB 1171 GCGCGGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1230
QY 1201 TCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
DB 1228 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1297
QY 1261 TTCAGGCGGCGGCTGTCGNNACAGATGAGAGAGTCTCGCTCGGCTCGGCTGATC 1320
DB 1288 TTCAGGCGGCGGCTGTCGNNACAGATGAGAGAGTCTCGCTCGGCTCGGCTGATC 1347
QY 1321 GGCAGGCTCACTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
DB 1348 GGCAGGCTCACTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1407
QY 1381 GCTTCGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1440
DB 1408 CTTGCTGCTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1467
QY 1441 GCGCGGCGGCTTCACTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1468 CATGCTGCTTCACTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1527
QY 1501 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
DB 1528 CAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1587
QY 1561 ACNGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
DB 1588 ACTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1647
QY 1621 CAGGCGGCTGAGCTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
DB 1648 CAGGCGGCTGAGCTTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1707

QY 1681 MCNNGCTGTCNAGAGCACTTGGCNCGCCCTTGAGCGCNNNNNNNNNNNACGAA 1740
Db 1708 CCGACTCTCTCCAGAGCACTCGGCACTGGCTTGAGCTC-----AACCA 1755
QY 1741 CTGNGGACAGAGTCAACAGNCGCTCNAAGCNCCTGAGCAGACCACTCTAGAC 1800
Db 1756 CTTCGCTCGAGGTCAAGAGGCGCTCAACAGCCTCTGAGCAGACAGCAGTACAC 1815
QY 1801 CTGAGCCGCGCTGAGCAGACGCTTCTGTTGCGAGCCGACCGCTGCGANNNN 1850
Db 1816 CTGAGCCGCGCTGAGCAGACGCTTCTGTTGCGAGCCGACCGCTGCGANNNN 1875
QY 1861 NNNGCTCTGCGCNCNNNAGAGTCTGCTGCGCNCGCTGAGAGCGTGAAGTGGCT 1920
Db 1876 TCGTCTCTGCGCTCTGCGCA--AGTCACTCTTACTGCGCTCAACGCTGGAAGTGGCT 1933
QY 1921 CCGCGGAGAGGCACTCTGCTGCTGAGCGGCGGANGTCCGAGCNCCTTGTGAGCGCCGT 1980
Db 1934 CCGCGGAGAGGCACTCTGCTGCTGAGCGGCGGANGTCCGAGCNCCTTGTGAGCGCCGT 1993
QY 1981 CCGTCTGCTGCGCGCGCTCNCCTGCTGCTGCGCGGCGGCGGCGCTGCTGCTG 2040
Db 1994 CTTCGAGGCGCGCGGCGGCGAGCATCTCTGCGCGGCGGCGGCTCTGTACTGTTG 2053
QY 2041 TCGCGGAGAGCTCGGCGCTCAAGCGCGCGGCGGCGGCACTTCTCTGCGGCAAGAGAG 2100
Db 2054 TGGCGGAGAGCTCGGCGCTCAAGCGCGCGGCGGCGGCACTTCTCTGCGGCAAGAGAG 2113
QY 2101 TGAAGTGGGCAACAGCTTCTCGGCTCTGAGAGCGCATCAAGNCGCGCATCAAC 2160
Db 2114 AGAGATCGGAGCAAGCTCTCGGCTCTGAGAGCGCATCAAGAGCGCGCATCAAC 2173
QY 2161 ACGTCTCGTCAAGATGCTCGCTAGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2220
Db 2174 ACGTCTCGTCAAGATGCTCGCTAGNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2232
QY 2221 NNNCCNNNNNNCTTNGNNNTGNNNTGNNNTGNNNTGNNNTGNNNTGNNNTGNNNTGNN 2280
Db 2233 ACCCAAGACAGCTTTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2292
QY 2281 TNNNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCTNNCT 2340
Db 2293 TGTGCTTACTCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2352
QY 2341 NGTNNCCANNNAACNCTNNNNNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 2400
Db 2353 GGTGCTCA--GTACAGCTGATAGAGCTGGAATGATGAGCTTTCGAGTTCAAAAAA 2411
QY 2401 NNNNAAAA 2408
Db 2412 AAAAAAAA 2419

RESULT 10
AAD45812
ID AAD45812 standard; DNA; 2419 BP.
XX
AC AAD45812;
DT 27-DEC-2002 (first entry)
XX
DE Rhodotorula graminis PAL DNA #1.
XX
KM Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KM cancer; human immunodeficiency virus infection; HIV; gene therapy; hcmv;
KM human cytomegalovirus infection; cytostatic; virocid; gene; ds.
XX
OS Rhodotorula graminis.
XX
FH Key Location/Qualifiers
FT CDS 37..2199
FT /*tag= a

FT /product= "PAL protein"
FT /EC_number= "4.3.1.5"
FT /transl_except= (pos:493..495, aa:Xaa)
FT /note= "Xaa corresponds to Val or Ala"
US2002102712-A1.
01-AUG-2002.
24-AUG-2001; 2001US-00939408.
24-JUL-2000; 2000US-00624693.
24-JUL-2001; 2001WO-US023270.
PCBU SERVICES INC.
Yoshida RK, Koester AB;
WPI; 2002-690616/74.
P-PDB; AAE27938.
Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
polypeptide, useful for treating a mammal having phenylketonuria, cancer,
human immunodeficiency virus or human cytomegalovirus infection.
Claim 13; Page 27-29; 74pp; English.
The present invention relates to Yeast (e.g. Rhodotorula) phenylalanine
ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
such proteins. PAL sequences are useful for producing L-phenylalanine.
They are useful for treating mammals having diseases, disorders or
conditions that would benefit from treatment with PAL proteins such as
phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
human cytomegalovirus (hcmv) infection. Sequences of the invention are
also used in gene therapy. The present sequence is R. graminis PAL DNA
Sequence 2419 BP; 416 A; 909 G; 659 C; 434 T; 0 U; 1 Other;
Query Match 65.8%; Score 1629; DB 6; Length 2419;
Best Local Similarity 78.0%; Pred. No. 9.1e-219;
Matches 1878; Conservative 1; Mismatches 504; Indels 25; Gaps 8;
QY 1 ATGGCCCCCTGCTGAGTCTGATCGACCTCTGTCGCGGCAAGCGGCTCNCNAGATG 60
Db 37 ATGGCCCCCTGCTGAGTCTGATCGACCTCTGTCGCGGCAAGCGGCTTACCAAGGCTG 96
QY 61 CAGCGCCGNNCCGNCNNNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 120
Db 97 CAGCGCGCTCGGACCAAGTGGGCTGGGCGGCCCACTTGGGCTGCGCGGCAAGC--CG 154
QY 121 GCTGCTCTCTCCGACCAACCAAGNAGAGCTGACATCTGTAAGNAATCTTGGCG 180
Db 155 GCTGATGAGCAGCGCGCGGCAAGTGCAGCTGAGATCTGCAAGAGCTCTCAAGC 214
QY 181 ACCCAAGCAGGACGNNNTGCAACTGCAAGGCTGACACCTTCACTCGAGAGCTG 240
Db 215 ACCCAAGCAGC---GACGCTGCTGAGCTCAGCGGCTGACAGCTTCAAGCTTCCGAGCTG 271
QY 241 TGGGCGCGGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
Db 272 TGGGCGCGGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 330
QY 301 AAGATCGACAAAGAGTCTGAGTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 331 CCGGTCGACAAAGAGTCTGAGTCTCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
QY 361 ACGATGCTGTTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
Db 391 ACCAGGCTTTCGGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450
QY 421 GCGTCTCTGAGCAGCACTCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
Db 451 GCGTCTCTGAGCAGCACTCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 510

QY 481 GAGCGGCGCTTCGAGAACTCGCTTCGCTCGAGGTGCTCCGCGCGCATGACCATTCGC 540
 DB 511 GAGCGGCGCTTCGAGAAACAGCTTCGCTCGAGGTGCTCCGCGCGCATGATATCCGC 570
 QY 541 GTCAACTCGCTCAGCAGCGGCGCACTCGGCGNCGCGCTCGTGCCTCGAGAGCGCTCACC 600
 DB 571 GTCAACTCGCTCAGCAGCGGCGCACTCGGCGNCGCGCTCGTGCCTCGAGAGCGCTCACC 630
 QY 601 AACTTCCTCAACACAGCGCATCACCCCATCGTCCCCCTCGGCGGACCATTCGCGCTCG 660
 DB 631 AACTTCCTCAACACAGCGCATCACCCCATCGTCCCCCTCGGCGGACCATTCGCGCTCG 690
 QY 661 GCGGACCTTCCTCCGCTCCTCCTCACTCGCGCGCGCATCACCGCTCAACCGGACNNCAAG 720
 DB 691 GCGGACCTTCAGCCCGCTTCGTAATCGCGCGCATCAACCGGTCAACCCGAGCTCAAG 750
 QY 721 GTNCAAGTNNNTNACAGAGGCGGCGANNAGAAGATCATGTNCGCCGCGAGCGATCGCGCTC 780
 DB 751 GTTCAAGTTTTCAGAGGGAACGAGAAGATCATGTTTTCGCGCGCGAGGCGCATTCGCTC 810
 QY 781 TTNGGTCTGAGACCCGCTGCTCTCGGCCCGAAGAGAGGTCTCGGTCTCTGTAACAGCAGC 840
 DB 811 TTGAGCTTCGAGAGAGGTGCTCTCGGCCCGAAGAGAGGTCTCGGTCTGTAACAGCAGC 870
 QY 841 GCGGTCTCCGCTCGATGAGCGACCTCGCTCTGCAAGACGACACATGCTCTCGCTCTC 900
 DB 871 GCGGTCTCCGCTCGATGAGCGACCTCGCTCTGCAAGACGACACATGCTCTCGCTCTC 930
 QY 901 TCGCAGGCGCTCAAGGCTCTMAAGGTCAAGGTCAAGGTCTGAGCGACGCGCGCTCTTCCAC 960
 DB 931 TCGCAGGCGCTTCAAGGCTCTCAAGGTCAAGGTCAAGGTCTGAGCGACGAGGCTCTTCCAC 990
 QY 961 CCGTTCCTTCAGAGAGTCAAGGCGCGCTCAACCGGACCGAGATCGAGGTCTGAGCGCGCAATC 1020
 DB 991 CCGTTCATCAAGAGGTCTGAGCGCGCGCTCAACCGGCGCGAGGTCTGAGCGCGCAATC 1050
 QY 1021 CCGCAGCTCTCGAGAGGCGACGNNGTTTTCGCTCAACAAGAGAGAGGTCAAGGTCAAG 1080
 DB 1051 CCGCAGCTCTTTCGCGGCTCGTGGTTTCCGTTAGACAGAGAGAGGTCAAGGTCAAG 1110
 QY 1081 GAGCAGAGAGGCAATTCCTCGCGCAGAGCGCTTACCCTTCGCGAGTGGCTCTCAATGCTC 1140
 DB 1111 GAGCAGAGAGGCAATTCCTCGCGCAGAGCGCTTACCCTTCGCGAGTGGCTCTCAATGCTC 1170
 QY 1141 GCGCGCGCTCGTCAAGCATGATTCACGCGCCACGCGNGTCTCTCGCTCGAGAGCGGAGAG 1200
 DB 1171 GCGCGCGCTCGTCAAGCATGATTCACGCGCCACGCGNGTCTCTCGCTCGAGAGCGGAGAG 1227
 QY 1201 TCGACGACCGGACACCCGCTCATCGAGCTCGAGAACAAAGANGACCCACACGCGCGCAAC 1260
 DB 1228 ACAGACGACCGGACACCCGCTCATCGAGCTCGAGAACAAAGANGACCCACACGCGCGCAAC 1287
 QY 1261 TTCCAGAGCGNCCGCTGCTCGCMAACAGATGAGAGAGCTCGCTCGCNCCTGCGCTGATC 1320
 DB 1288 TTCCAGAGCGNCCGCTGCTCGCMAACAGATGAGAGAGCTCGCTCGCNCCTGCGCTGATC 1347
 QY 1321 GCGCAGCTCAACTTCACGAGCTCAACGAGATGCTCAACGCGCGCATGMAACGCGGCGCTN 1380
 DB 1348 GCGCAGCTCAACTTCACGAGCTCAACGAGATGCTCAACGCGCGCATGMAACGCGGCGCTN 1407
 QY 1381 CCGTTCCTGCTCGCTCGCGAGAGACCGTCTCTCTCTCACTGCAAGAGGCTCTGACATT 1440
 DB 1408 CCGTTCCTGCTCGCTCGCGAGAGACCGTCTCTCTCTCACTGCAAGAGGCTCTGACATT 1467
 QY 1441 GCGCGAGCGNCTACCTTCGAGAGCTCGGACCTTGCGCAACCGCGGTACAGACCAAGTCTC 1500
 DB 1468 CACATGCTGCTTACGCTTCGAGAGCTCGGCGCACCTTGCGCAACCGCGGTACATCACTTCTC 1527
 QY 1501 CAGCGCGAGAGATGAGGCAACAGGCGCGTCAACTGCTCGCNCCTCATCTCGGCGCGCGC 1560
 DB 1528 CAGCGCGAGAGATGAGGCAACAGGCGCGTCAACTGCTCGCNCCTCATCTCGGCGCGCGC 1587

QY 1561 ACNCGCAGAGCGCAAGAGCGTCTTCTCTCCTCCTCGGCAACCAACCTCTCATGCGTNTCTC 1620
 DB 1588 ACTGCGAGAGCGCAAGAGCGTCTTCTCTCTCTCTGCTCGACCTCTCTGACGCTC 1647
 QY 1621 CAGGCGCTGACCTCCGCGCATGAGATTGATTAAGAAAGATTGACACCGTNNCTC 1680
 DB 1648 CAGGCGCTGACCTCCGCGCATGAGATTGATTAAGAAAGATTGACACCGTNNCTC 1707
 QY 1681 NCGNCGCTCCTNAGCAGCACTTTGGCNCNCCCTCGACCGCNCNNNNNNNNNNAAGAA 1740
 DB 1708 CCGACTCTCTCCAGCAGCACTCGGCACTGGCTCGACGCTC-----AAGCA 1755
 QY 1741 CTGCGGACAAGAGTCAACAAAGCCTCCTNACAGAGCNCCTCGAGACCAACCTGTAAGAC 1800
 DB 1756 CTGCGGCTCGAGGTCAAGAAAGCGCTCAACAGGCTCTCGAGAGAGAGAGAGTACGAC 1815
 QY 1801 CTGAGCCGCGCTGACAGACCGCTTCTGTTGCGACACCGGACCGTCTGCAAGNNNNNN 1860
 DB 1816 CTGAGCCGCGCTGACAGACCGCTTCTGTAACCGACCGGACCGCTGAGGCTCTC 1875
 QY 1861 NNGTCTCGCCNNNNGCCANNAGGTCTGCTCGGCGCGTCAACGCTTGAAAGTGCCT 1920
 DB 1876 TGTCTCTCGCTCTGCGCA--ACGTCAACCTTACTGCGCTCAAGCGCTGAAAGTTCCT 1933
 QY 1921 CCGCGGAGAAGCCATCTCGCTCAAGCGCNAAGTCCGCAACCTTCTGANCAGCNCCT 1980
 DB 1934 CCGCGGAGAAGCCATCTCGCTCAAGCGCGAGGCGCAACCGCTTCTGAGACAGCGCT 1993
 QY 1981 CGTCTCGTCCCGCGCTGCMCGTACTCTGCGCGCGCAACGCGGTCTGTAACGCTTTC 2040
 DB 1994 CTTGCGAGCGCGCGCGCAAGCATACCTTCCGCGCGCAACGCGGTCTGTAACGCTTTC 2053
 QY 2041 TCGCGAGAGCTCGGCTCAAGGCGCGCGCGCGCGAGCTTCTCTCGGCAAGAGAGAG 2100
 DB 2054 TCGCGAGAGAGCTCGGCTGAGGCGCGCGCGCGCGCGAGCTTCTCTCGGCAAGAGAG 2113
 QY 2101 TGACGATGCGCACCAAGCTCTCCGCACTTACGAGGCTTAAAGNNGGCNCATCAAC 2160
 DB 2114 AGACCATCGGAGGCAAGCTCTCGGCACTTACGAGGCTTAAAGAGGCGCGCATCAAC 2173
 QY 2161 AGTCTCGTCAAGTGTGCGTAAAGNNCNCNNCNCNNCNCNNCNCNNCNC 2220
 DB 2174 AGTCTCGTCAAGTGTGCGTAAAGNNCNCNNCNCNNCNCNNCNCNNCNC 2232
 QY 2221 NNNCNCNNNNNNCTTNGANNCTGANNCTGANNCTGANNCTGANNCTGANNCTGANN 2280
 DB 2233 ACCCGAAGCAGCTTTCGAGGTGCTGCGTCTGCGCAAGACGAGCTTCTCATACACA 2292
 QY 2281 TNNNNCTNNCTNNCTCNCNNNNNCCNNCTCNCNNNNCTNNNNCTNNNNCTNNNN 2340
 DB 2293 TGTGCGCTTACTCTCTGCGCGCTCATCAGCTCTCTCACTTCTGTAACCGCGTCTC 2352
 QY 2341 NGTNNCANNACNTNNNNNNNANNNGANNNGANTNNANGNTNCGANNNNNNNN 2400
 DB 2353 GGTGCTCA-GTACAGCTGTATAGAGCCTGGAATGATTGCAAGCTTCAAAAAA 2411
 QY 2401 NNNANAAA 2408
 DB 2412 AAAAAAAA 2419

RESULT 11
 AAD3076
 ID AAD3076 standard; DNA; 2151 BP.
 AC AAD3076;
 XX
 XX
 DT 01-JUL-2002 (first entry)
 XX
 XX Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant DNA.
 XX Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;

KW cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
 LW LCP; telecommunication; medical device; aerospace application; enzyme;
 KM biocatalyst; gene; mutant; variant; ds.
 XX Rhodotorula glutinis.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT 1..2151
 FT CDS /product= "Rhodotorula glutinis mutant PAL"
 W0200210407-A1.
 PD 07-FEB-2002.
 XX
 PF 23-JAN-2001; 2001WO-US002099.
 XX
 PR 27-JUL-2000; 2000US-00627216.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Gatebny AA, Sarielani FS, Tang X, Qi WW, Vannelli T;
 XX
 DR WPI; 2002-280635/32.
 DR P-PSDB; AAE20654.
 XX
 PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT polypeptide, or mutant TAL polypeptide, useful for producing
 PT parahydroxycinnamic acid in recombinant host cell lacking cinamamate
 PT hydroxylase.
 XX
 PS Example 10; Page 95-98; 139pp; English.
 XX
 CC The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
 CC plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
 CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
 CC recombinant techniques. The recombinantly produced PHCA may be used as a
 CC monomer for production of liquid crystal polymers (LCP). LCP may be used
 CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC PAL mutant DNA
 CC
 XX
 SO Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;

Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 2.1e-218;
 Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGGACCCCTCGATCGATCGAGCTCGTGCAGAGGCGTNCNAAACGATNG 60
 DB 1 ATGGACCCCTCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 60
 QY 61 CACGCCGNNCCGNNNGCNCACGGGCGCCACGTCACNCTTCNGCGCCGNCNG 120
 DB 61 CAGGCTGT-----CAATGCGCTTCGACCAACTTC-----GCACTCCAG 100
 QY 121 GCTGCTCTCCGACACCCAGNNGACGAGCTCGACATGCTGAGNAGATCTCCCG 180
 DB 101 GCTGACCTGCTCCACACCCAGGTCACGAGTGCATGCTGAGAGATGCTCCCG 160
 QY 181 ACCCCACCGNNAAGNAGNNNTGAACTTCGACGGGTACACCTTCAGGAGAGCTG 240
 DB 181 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 161 CGCCACGCACTCGA---CGCTCGAATCGACGGCTACTCGCTCAACCTCGAGAGCTCG 217
 QY 241 TCGGCGCGCGCNCGCAAGGCGCGCNCNGTCCGCTTCNACAGACAGNCGACGAGATCCGCGCA 300
 DB 218 TCTCGCGCGCGAGAGGCGAGGCGCTGCTCCCGCTCAAGAGACAG-CGACGAGATCCGCTCA 276
 QY 301 AAGATCGACAANAGNTGAGTTCCTCCGNNCNCAGCTTCNACACAGAGTCTACGAGTGC 360
 DB 277 AAGATTGACAAATTCGATTGAGTTCTTGCGCTCGAATCTTCATGACGCTTCAGGCGTTC 336
 QY 361 ACGACTGATTTCGCGGCTCGGCGCGACACCCGACGTAAGATGCAATCTGCTCCAGAG 420
 DB 337 ACGACTGGAATTGGGGATTCGAGACACCGGACCGAGGACGGCATCTGCTCCAGAG 396
 QY 421 GGNCTCTCGACACACCACTCTGGGGTGTNTCTCCNACGTCGANTCGANTCTTCNGCTTC 480
 DB 397 GCTCTCTCGAGACACCACTCTGGGGTGTNTCTCCNACGTCGANTCGANTCTTCNGCTTC 456
 QY 481 GGNCCGGCTTCGAGAACTCGCTTCGCTGAGGTCGTCGCGGCGCATGACCATCCG 540
 DB 457 GGGCGCGCTTCGAGAACTCGCTTCGCTGAGGTCGTCGCGGCGCATGACCATCCG 516
 QY 541 GTCAACTGCTCACNCGCGCCACTCGGCGNGTCGCTCTGCTCTCGAGGCGCTACC 600
 DB 517 GTCAACAGCTTGACCCCGGCGCACTCGGCTGTCCGCTCTGCTCTCGAGGCGCTACC 576
 QY 601 AACTTCCTCAACCAAGGATCACCCCATCGTCCGCTCCGCGGACATCTGCGGCGTG 660
 DB 577 AACTTCCTCAACCAAGGATCACCCCATCGTCCGCTCCGCGGACATCTGCGGCGTG 636
 QY 661 GCGACCTCTCCGCT 720
 DB 637 GCGACCT 696
 QY 721 GTCACGNNNNNACGAGGAGGANNAGATATGNNCCCGCGGAGGATCGGCGTC 780
 DB 697 GTGACGCTGTTCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
 QY 781 TTGAGTCTCGAGCCGCTGCTCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
 DB 757 TTCAACTCTGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
 QY 841 GCGTCTCGGCTGATGCGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 DB 817 GCGGCTCGAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 876
 QY 901 TCGAGGCGCTCAAGGCTCTTACGTCGAGGCGATGTCGCGGCGGCGGCTGCTGCTGCTG 960
 DB 877 TCGAGTGTCTCAAGGCGCTTACGTCGAGGCGATGTCGCGGCGGCGGCTGCTGCTGCTG 936
 QY 961 CCGTTCCTCAAGAGTCAGCGGCTTACGTCGAGGCGATGTCGCGGCGGCGGCTGCTGCTGCT 1020
 DB 937 CCGTTCCTCAAGAGTCAGCGGCTTACGTCGAGGCGATGTCGCGGCGGCGGCTGCTGCTGCT 996
 QY 1021 CGGACGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
 DB 997 CGGAGGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
 QY 1081 GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
 DB 1057 GACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
 QY 1141 GGGCGCTCTGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
 DB 1117 GGGCGCTCTGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
 QY 1201 TCGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
 DB 1177 TCGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1236
 QY 1261 TTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
 DB 1237 TTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1296

QY	1321	GGCAAGCTCAATTACGACACTCACCGAATGCTCAACGCGGATGAACGCGGCTT	1380
Db	1297	GGAAAGCTAAATTACGACACTCAAGATGCTCAACGCGGATGAACGCGGCTT	1356
QY	1381	CCNTCTGCTCGCTGCGGAGAACCTCGCTCTCTTATCTATGCAAGGGCTTGACATT	1440
Db	1357	CCCTCTGCTCGCGGGCGAAGACCCCTCGCTCTCTTACACTGCAAGGGCTTGACATT	1416
QY	1441	GGCGGCGGCTTACCTTGGAGCTGGGACCTTGGCAACCGCGTACAGCCACGTC	1500
Db	1417	GGCGCTGGGGGTACCTTGGAGCTGGGACCTTGGCAACCTTGTGACGACGATGTC	1476
QY	1501	CAGCGCGGAGATGGGCAACAGGCGGTAACTCGCTGCMCTCATCTCGGCGCGCG	1560
Db	1477	CAGCGCGGCTGAGATGGGCAACAGGCGGTAACTCGCTTGGCTCATCTCGGCTGTCG	1536
QY	1561	ACNGCGGAGCCAAAGAGTCTTCTTCTCTCTCTGCGCACCTCTACTGCGTNTCT	1620
Db	1537	ACACCGGAGTCCAAAGAGTCTTCTTCTCTCTGCGCACCTCTACTGCGTNTCT	1596
QY	1621	CAGCGCGGTGACCTCGCGCATGAGTTCGAGTTGACAGAGAGTTGACCCGATNTTC	1680
Db	1597	CAAGCATGACTTTCGCGGACCGAGTTCGAGTTGACAGAGAGTTGCGGCCACATTC	1656
QY	1681	NCNCGCTCTNACGACACTTTGGGCMCGCTCGACGCGCMNNNNNNNNNATACAA	1740
Db	1657	GTCTGCTATGACACGACACTTTGGGCTCGCATGACGCGCTCGAACCTGCGCACAG	1716
QY	1741	CTCGNNGACAAGTCAACAAGNCCTNACAAAGGNCCTCGAGAGACCACTGTACAC	1800
Db	1717	CTCGTGAAGAAGTCAACAAGCCTGCGCAAGGCTCGAGAGACCACTGTACAC	1776
QY	1801	CTGAGCCGCGCTGGCAAGACGCTTCTGTTGCGCACCGGACCGTGTGAGNNNNN	1860
Db	1777	CTGCTGCCGCTGGCAAGACGCTTCTGTTGCGCGCGGACCGTGTGAGGTCCTC	1836
QY	1861	NGTCTCTGCMNNNGCCANNAGTCTCGCTCGCGGCGCTCAACGCTGAGAGTCTGCT	1920
Db	1837	TGCTGACGCTC-----GCTCTGCTCGCGCGGCTCAACGCTGAGAGTCTGCTG	1885
QY	1921	CGCGCGAGAGGCAATCTGCTCACGCGCMANGTTCGCGACNCCTTTCTGANCNCCT	1980
Db	1886	CGCGCGAGTGGCCATCTGCTCACCGCGCAAGTTCGCGAGACCTTCTGCTCGCGGCT	1945
QY	1981	CGTGTGCTGCGCGCGCTGCMGTACCTCTGCGCGGACGCGGCTGCTGACTGTTG	2040
Db	1946	CGACTCTGCTGCGCGGCTCTGTAACCTCTGCGCGGACTGAGATCTCTAGCGCTTC	2005
QY	2041	TCCGCGAGAGCTCGGCTCAAGGCGCGCGCGCGGACGCTTCTTCTGCGGACGAGG	2100
Db	2006	TCCGCGAGAGCTTGGCTCAAGGCGCGCGCGGAGACGCTTCTTCTGCGGACGAGG	2065
QY	2101	TGACGATCGGACCAAGCTCTCCGACCTTACAGAGGACATCAAGNNCGCMCATCAAC	2160
Db	2066	TGACGATCGGCTCGAAGCTTCTTCAAGATCTACAGGCGCATCAAGTGGGAGAGTCAACA	2125
QY	2161	ACGTCTCTGTCAAGTGTCTGCGTAA	2186
Db	2126	ACGTCTCTCTCAAGATGCTGCTTAA	2151
RESULT 12			
AAD33100 standard; DNA; 2151 BP.			
XX	AAD33100;		
XX	AAD33100;		
XX	01-JUL-2002 (first entry)		
XX	R. glutinis phenylalanine ammonia lyase (PXL) mutant DNA, Epi8Km-6.		
XX	Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;		
XX			

XX	phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KW	cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KM	LCP; telecommunication; medical device; aerospace application; enzyme;
XX	biocatalyst; mutant; variant; gene; ds.
OS	Rhodotorula glutinis.
XX	Synthetic.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2151
FT	/tag= a
FT	/product= "Rhodotorula glutinis PAL mutant"
FT	/EC_number= "4.3.1.5"
FT	mutation
FT	/replace(645, G)
FT	/tag= b
FT	mutation
FT	/replace(792, A)
FT	/tag= c
FT	mutation
FT	/replace(858, T)
FT	/tag= d
FT	mutation
FT	/replace(1619, T)
FT	/tag= e
XX	MO200210407-AI.
XX	07-FEB-2002.
PF	23-JAN-2001; 2001WO-US002099.
PR	27-JUL-2000; 2000US-00627216.
XX	(DUPO) DU PONT DE NEMOURS & CO E I.
PA	Gatenby AA, Sarasiani FS, Tang X, Qi WW, Vannelli T,
P1	WPI: 2002-280635/32.
DR	P-PSDB; AAEZ0668.
PT	Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT	polypeptide, or mutant TAL polypeptide, useful for producing
PT	parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT	hydroxylase.
XX	Disclosure; Page; 139pp; English.
XX	The invention relates to methods for biological production of para-
XX	hydroxycinnamic acid (PHCA). The invention relates to developing of a new
CC	biocatalyst for conversion of glucose to PHCA by incorporation of the
CC	wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
CC	Rhodotorula glutinis into Escherichia coli underlying the ability of the
CC	wildtype PAL to convert tyrosine to PHCA. The invention is also directed
CC	to developing a new biocatalyst for conversion of glucose to PHCA by
CC	incorporation of the wildtype PAL which possesses enhanced tyrosine
CC	ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
CC	plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
CC	Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
CC	recombinant techniques. The recombinantly produced PHCA may be used as a
CC	monomer for production of liquid crystal polymers (LCP). LCP may be used
CC	in electronic connectors and telecommunication and aerospace
CC	applications. LCP resistance to sterilizing radiation has also enabled
CC	these materials to be used in medical devices as well as chemical, and
CC	food packaging applications. The present sequence is Rhodotorula glutinis
CC	PAL mutant DNA. Note: This sequence is not shown in the specification,
CC	however it is constructed based on the PAL DNA shown as SEQ.ID.NO:7
CC	(AABJ3075) in the sequence listing
XX	Sequence 2151 BP; 382 A; 830 C; 565 G; 374 T; 0 U; 0 Other;
XX	Query Match 65.7%; Score 1626.4; DB 6; Length 2151;
XX	Best Local Similarity 83.7%; Pred.No. 2.le-218;
XX	Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;
XX	1 ATGGCCCCCTCGATCGATCGATCGACACTGCTGTCGACAGCGCCTGCNMAAGGANNING 60

```
Db 1 ATGGACACCTCGCTCGATCGATCTTGGCACTTGGCTTGGCAAGCGGCTGCAATCCGCAAG 60
QY 61 CACGCCGNNCCGNCNNNGNCNMACGGGCGCCACGTCACNCTGNCNGCCGNCNG 120
Db 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GAGTCGAG 100
QY 121 GCTCGCTCTCCCGGACCAACGANNACCGCACTGACATCGTTNAGNAGTCTCTCGCG 180
Db 101 GCTCGACCTGCCCAACCCAGTCAAGAGTGAATGCTGAGAGATGCTCGCG 160
QY 181 ACCCAACCGNNACGNAAGNNNTGAACTGACGGGTAGACCCCTACCTCGANAGCTCG 240
Db 161 CGCCGACCGACTGA---GGCTGAACTGACGGCTACTCGTCAACCTCGAGACGTCG 217
QY 241 TCGGCGCCGNCGCAAGGGCCGNCNGTCCGCTGNCAGACAGNCAGAGATCCGCGCA 300
Db 218 TCTCGGCGCGGAGAGGAGGAGGCGCTGTCCGCTGAGAGAGAG-AGACGAGATCCGCTCA 276
QY 301 AAGATCGACAAAGNGTGGAGTTCTCCGNNCNCAGCTCNACACAGNGTGTACGAGTGC 360
Db 277 AAGATTGACAAATCGGTCGAGTTCTTGGCTCGCAACTCTCATAGAGCTTACGGGTC 336
QY 361 ACGACTGTTTCGGGCGCTCGGCGCAACCCGAGTGAAGATGCAATCTCGCTCAGAG 420
Db 337 ACGACTGGAATTGGCGGATCCGAGACACCCGCAACGAGAGAGCCATCTCGCTCAGAG 396
QY 421 GNCCTCTCGAGACACAGCTTCGGGTGNNCTCCNMACGTCGNTGANTCTTNGCTC 480
Db 397 GCTCTCTCGAGACACAGCTTCGGGTGTTCTCTCTGCTGATGACTCTGTCGCTC 456
QY 481 GNCGCGGCGCTCGAGACTCGCTTCGCTCGAGTGTGTCGCGCGGCGCCATGACATCGC 540
Db 457 GCGCGGAGTCTCGAGAACTCGCTTCCCTCGAGGTGTTTCGGGCGCCATGACAAATCGC 516
QY 541 GTAACTCGCTCAGNCGGCGCACTCGGCGNGTCCGCTGCTGTCCTGAGGCGCTCAC 600
Db 517 GTAAAGCTTGACCCGCGCCACTCGGCTGTCCGCTGTCGCTCGAGGCGCTCAC 576
QY 601 AACTTCTCAACACAGGAGTCAACCCCATGTCCTCCGCGGACCATCTCGGCGTCG 660
Db 577 AACTTCTCAACACAGGAGTCAACCCCATGTCCTCCGCGGACCATCTCGGCGTCG 636
QY 661 GCGGACCTCTCCCNCTCTCTNATCATGCGCGCGCCATCAACGGGTACCCGAGCANNAG 720
Db 637 GCGGACCTCTCTCTCTCTCTACTGAGAGGCGCATCAAGGATCCCGACAGAG 696
QY 721 GTNCAAGTNNTNCAGAGGCGANNAGAAAGATCATGTCGCGCGGCGAGCGATCGGCTC 780
Db 697 GTGACAGTGTCAAGAGGCGAAAGAGAGATCTGTAGCGCCCGAGGCGATGCGCTC 756
QY 781 TTNGTCTCGAGCCGCTCTCTCGGCGCGGAGAGAGGTCGCTGTCACAGGACG 840
Db 757 TTGAACCTGAGCGCGTCTCTCGGCGCGGAGAGGTCGCTGTCACAGGACG 816
QY 841 GCGCTCTCGGCTCTGATGCGGACCTCGCTCTGACAGAGCAACATGCTCTCGCTCTC 900
Db 817 GCGCTCTGAGCATGATGCGGACCTCGCTCTGACAGAGCAACATGCTCTCGCTCTC 876
QY 901 TCGAGAGGCGCTCAAGGCTCTNACGGTTCAGAGGCGATGTCGCGCAAGCGCGGCTGTTCCAC 960
Db 877 TCGAGTCTCTCAAGGCGCATGACGGTTCAGAGGCGATGTCGCGCAAGCGCGGCTGTTCCAC 936
QY 961 CCNTTCTCTCAAGAGTCAAGCGGCTCAACCCGACCAAGATGAGGTCGCGCGCAATC 1020
Db 937 CCCTTCTCTCAAGAGTCAAGCGGCTCAACCCGACCAAGATGAGGTCGCGCGCAATC 996
QY 1021 CGAGGCTCTCGAGGCGAGCANNGTTCGCGGCAACCGAGAGGAGGTCAAGGTCAAG 1080
Db 997 CGGAGCTCTCGAGGAGCGCGCTTGTGTGTCACATGAGAGAGAGGTCAAGGTCAAG 1056
QY 1081 GACGAGAGAGGATTTCTCGGAGAGACCGCTACCGGCTCGCACGTCGCTCACTGAGTGC 1140
Db 1057 GACGAGAGAGGATTTCTCGGAGAGACCGCTACCGGCTCGCACGTCGCTCACTGAGTGC 1116
QY 1141 GGCCCGCTGTGTCAGGACATGATTTACGCCCAAGCNCNTCTCTCTGCTCGAGGCGGNCAG 1200
Db 1117 GGCCCGCTGTGTCAGGACATGATTTACGCCCAAGCNCNTCTCTCTGACATCGAGGCGGNCAG 1176
QY 1201 TCGAGACCGACCAACCCGCTCATGAGCTGAGAGAACAGAGAACCCACGACGCGGCGAC 1260
Db 1177 TCGAGACCGACCAACCCGCTCATGAGCTGAGAGAACAGACTTGTGACACAGGCGGACAT 1236
QY 1261 TTCAGAGGAGCGCTGTGAGAACAGATGAGAGAGACTGCGCTCGCNCCTCGGCTGATC 1320
Db 1237 TTCAGAGGAGCGGCTGTGAGAACAGATGAGAGAGACTGCGCTCGGAGCTCGCCAGATC 1296
QY 1321 GCGAAGCTCAACTTCAAGGACCTCAAGGATGTCACAGCGCGGATGAAACCGCGGCTC 1380
Db 1297 GCGAAGCTCAACTTCAAGGACCTCAAGGATGTCACAGCGCGGATGAAACCGCGGCTC 1356
QY 1381 CCNTCTGCTCGCTGTCGAGAGACCCNTGCTCTCTATCATGCAAGGCGCTCGACAT 1440
Db 1357 CCCTCTGCTCTGCGGCGGAGAGACCCCTCGCTCTCTACACATGCAAGGCGCTCGACATC 1416
QY 1441 GCGCGGCGGAGTACCTTTCGAGCTGAGACCTTTCGCAACCCGCTGACACCTGTC 1500
Db 1417 GCGCGTGGGAGTACCTTTCGAGCTGAGACCTTTCGCAACCCGCTGACACCTGTC 1476
QY 1501 CAGCGGCGGAGATGAGGAGACAGGCGGTCACCTCGCTGCGNCATCTCGGCGGCGG 1560
Db 1477 CAGCGGCGGAGATGAGGAGACAGGCGGTCACCTCGCTGCGNCATCTCGGCGGCGG 1536
QY 1561 ACNCGGAGGCGCAAGAGCTCTTCTCTCTCTCTGCGCAACCACTCTACCTGAGTCTC 1620
Db 1537 ACGACGAGTCAAGAGAGCTCTTCTCTCTCTCTCTCTGCGCAACCACTCTACCTGAGTCTC 1596
QY 1621 CAGCGGCGGAGCTCTCGGCGGATGAGTGAAGTGAAGAGAGTTCGACCCGNNNTC 1680
Db 1597 CAGGCAATGAGCTTGGCGGACCGAGTGAAGTGAAGAGAGTTCGACCCGACATC 1656
QY 1681 NCCNCGCTCNCNAGACGACTTTGGCNCGCCCTCGACGCGNNNNNNNNNNNAGAA 1740
Db 1657 GTCTGCTCATGACAGGACATTTGGCTCCGCGCATGACGCGCTGAACTGCGGCGAGAG 1716
QY 1741 CTCGAGGACAAAGTTCACAAAGCGCTNACAAAGCNCNTGAGAGAGCACTGTGAGC 1800
Db 1717 CTCGAGGAGAAAGTTCACAAAGCGCTGCAAGGCGCTCGAGAGAGCACTGTGAGC 1776
QY 1801 CTCGAGCGGCGTGGCAAGAGCGCTTCTGTTGCGGACGCGGACCGCTGTCGAGNNNNN 1860
Db 1777 CTGTCGCGGCGTGGCAAGAGCGCTTCTGTTGCGGCGGACCGCTGTCGAGNNNNN 1836
QY 1861 NNCTCTCGCCNNNNGCCANNAGTCTGCTCGCNCGCTGCAAGCCTGAGAGTGCCT 1920
Db 1837 TCGTGAAGTC-----GCTCTGCTCGCGCGGTCACAGCCTGAGAGTGCCTG 1885
QY 1921 CCGCGAGAAAGGCTATCTGCTCAAGCGGANNATCGCGGACNCTTCTGNCGNCCT 1980
Db 1886 CCGCGAGATGAGGCTATCTGCTCAAGCGGAACTGCGGAGCTTCTGTCGCGCGCT 1945
QY 1981 CGTGCCTGCGCCGCGCTGNCGATCTCTGCGCGGACAGCGCGCTGTACTGCTTGC 2040
Db 1946 CGACTGCTGCGCGCGCTCTGATCTCTGCGGAGATGATGATCTCTGAGCTTGC 2005
QY 2041 TCCGAGAGAGCTGAGGTCAGAGCGCGCGCGGAGAGCTTCTCTGCGGAGAGAG 2100
Db 2006 TCCGAGAGAGCTGAGGTCAGAGCGCGCGGAGAGAGCTTCTCTGCGGAGAGAG 2065
QY 2101 TGAAGATGAGCAAGCTCTCCGCACTTACAGAGGCGATCAAGNNCGCNCGATCAAC 2160
Db 2066 TGAAGATGAGCTCAAGCTCTCAAGATCTACAGAGGCTCAAGTGGGCGAGATCA 2125
QY 2161 AGCTCTCTGTCAGATCTCGCCTAG 2186
Db 2126 AGCTCTCTCTCAAGATCTCGCTAG 2151
```


QY 1321 GGCAAGCTCAACTTCACCGAGCTGACCGAGATGCTCAACGCCGCGATGAACCGCGGCTTN 1380
DB 1297 GGCAAGCTCAACTTCACCGAGCTGACCGAGATGCTCAACGCCGCGATGAACCGCGGCTTC 1356
QY 1381 CCNCTCTGCTCGCTGCCAGAGACCCTGCTCTCTATCACTGCAAGGGCTTCGACATT 1440
DB 1357 CCTCTCTGCTCGCGCGGCGAAGACCCCTCGCTCTCTACGACTGCAAGGGCTTCGACATC 1416
QY 1441 GCGCGNCGNCTTAACCTTCGAGCTCGGACCTTGCCAAACCGGTNAGACCCAGCTC 1500
DB 1417 GCGGCTGCGGCGTACCTCGAGTGGAGACCTTCGCCAACCTGTGACGAGCATGTC 1476
QY 1501 CAGCGGCGNAGATGGAGCAACGAGGCGCTCAACTGCTCGCCTCATCTCGGCGCGC 1560
DB 1477 CAGCGGCGTGANATGGAGCAACGAGGCGTCAACTGCTTGCGCTCATCTCGGCTGTCG 1536
QY 1561 ACGGCCGAGGCCCAAGAGTCTTTCTCTCTCTCTCGCCACCACTCTACTGAGTCTC 1620
DB 1537 ACGACCGAGTCCACGAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
QY 1621 CAGGCGGTGACCTCGCGCGCGATGAGTTGAGTTCAAGAGCAGTTGACCCGCTNTTC 1680
DB 1597 CAAAGCATGACTTGGCGGCGGACCGAGTTCAAGTTCAAGAGCAGTTGCGCCGACCATC 1656
QY 1681 NCNCGCTCTNTGACGAGCACTTTGCGCNCNGCCTCGACGCGCNCNNNNNNNNNNACGAA 1740
DB 1657 GTCTGCTCATGACGAGCACTTTGGCTCCGCAATGACGGCTCGAACTTGCGCGAGAG 1716
QY 1741 CTCGAGAGAAAGTCAACAAGNCGCTCAACAAGCNCCTCGAGCAGACCACTGCTAGAC 1800
DB 1717 CTCGTGAGAAAGTGAACAAGCCTCGCAAGCCTCGAGCAGACCACTGCTAGAC 1776
QY 1801 CTCGAGCCGCGCTGAGCAGACGCTTCTGCTTCGCGGACCGGACCGCTGCTGAGNNNNN 1860
DB 1777 CTGCTCCCGCGCTGGAGCAGACGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1836
QY 1861 NNNGTCTCGCCNNNNGCCANNAGTCTGCTGCGCTGCGCCGTCAAGCTGGAAGTCTGCT 1920
DB 1837 TCGTCGACGCTC-----GCTCTGCTGCGCGCGCTCAAGCCTGGAAGTCTGCGC 1885
QY 1921 CCGCGGAGAAAGGCACTCTGCTCACGCGGANNGTGCGGACNCCCTTGAGNCGGNCCTG 1980
DB 1886 CCGCGGAGTGGGCACTCTGCTCAACCCGCAAGTCCGCGAGACCTTCTGTCGCGCGCT 1945
QY 1981 CGTGTGCTGCGCCGCGCTCNCGTACTCTGTGCGCGGCAAGCGGCTCTGTACTGCTTCG 2040
DB 1946 CGACTCTGTGCGCCGCGCTCTGTACTCTGCGCGGCACTCAGATCTCTACGCTTCG 2005
QY 2041 TCGCGAGAGAGCTCGGCTTCAAGGCCCGCGCGAGAGTCTTCTCTCGGCAAGAGAGG 2100
DB 2006 TCGCGAGAGAGCTTGGCGTCAAGGCCCGCGCGAGAGCTTCTCTCTCGGCAAGAGAGG 2065
QY 2101 TGACGATCGGCAACAAGTCTCCCGATCTACAGAGGCATCAAGNCGGCGCATCAACC 2160
DB 2066 TGACGATCGGCTGGAAGCTCTCCAGATCTACAGAGGCATCAAGTGGGCAAGATCAACA 2125
QY 2161 ACGTCTCTCTGTAAGATGCTGCGNTAG 2186
DB 2126 ACGTCTCTCTCAAGATGCTGCTTAG 2151

Search completed: September 11, 2004, 14:50:34
Job time : 1376.94 secs

Db 241 TCGGCGCCGNCGCAAGGCGCGCNCNGTCGCGCTCNCAGACAGNCGACGATCCGCGCA 300
Qy 301 AAGATCGACAAANAGNGTCGAGTTCCTCCGNNCAGCTCNAACAAGNGTCTACGAGNTC 360
Db 301 AAGATCGACAAANAGNGTCGAGTTCCTCCGNNCAGCTCNAACAAGNGTCTACGAGNTC 360
Qy 361 ACGACTGTTTCGCGGCTCGCGGACACCCGAGCTGAGAGATGNAATCTCGTCCAGAA 420
Db 361 ACGACTGTTTCGCGGCTCGCGGACACCCGAGCTGAGAGATGNAATCTCGTCCAGAA 420
Qy 421 GCGTCTCTGAGACACGAGCTCTGCGGTGTCCTCCNACGTGANTGANTCTTCGCTC 480
Db 421 GCGTCTCTGAGACACGAGCTCTGCGGTGTCCTCCNACGTGANTGANTCTTCGCTC 480
Qy 481 GCGGCGGCTCGAGAACTCGCTTCGCTCGAGGTGTCGCGGCGCATGACATCCGC 540
Db 481 GCGGCGGCTCGAGAACTCGCTTCGCTCGAGGTGTCGCGGCGCATGACATCCGC 540
Qy 541 GTCAACTCGCTCAACGCGGCGCATCGGCGGCTCGCTGCTGCTGAGGCGCTCAC 600
Db 541 GTCAACTCGCTCAACGCGGCGCATCGGCGGCTCGCTGCTGCTGAGGCGCTCAC 600
Qy 601 AACTTCTCAACACGAGCATCACCCCATCTGTCCTCCCTCGCGGACCATCTCGCGCTCG 660
Db 601 AACTTCTCAACACGAGCATCACCCCATCTGTCCTCCCTCGCGGACCATCTCGCGCTCG 660
Qy 661 GCGGAGCTCTCCCT 720
Db 661 GCGGAGCTCTCCCT 720
Qy 721 GTNACGCTNNTNACAGAGGCGCANNGAGAGATCATGTGCGCGGAGGAGCATCGCGCTC 780
Db 721 GTNACGCTNNTNACAGAGGCGCANNGAGAGATCATGTGCGCGGAGGAGCATCGCGCTC 780
Qy 781 TTNGGTTCGAGCCCGCTCGTCTCGGCGCGAAGAGAGGTCTTGCTTGTCTCAAGGACG 840
Db 781 TTNGGTTCGAGCCCGCTCGTCTCGGCGCGAAGAGAGGTCTTGCTTGTCTCAAGGACG 840
Qy 841 GCGGCTCTCGGCTCGATGAGGAGCCCTCGCTCTGACGACGACACATGCTCGCTCTC 900
Db 841 GCGGCTCTCGGCTCGATGAGGAGCCCTCGCTCTGACGACGACACATGCTCGCTCTC 900
Qy 901 TCGAGAGCGCTCAAGGCGCTTACGAGTCAGGCGCATGCTGAGCGCGCTCGTTCCAC 960
Db 901 TCGAGAGCGCTCAAGGCGCTTACGAGTCAGGCGCATGCTGAGCGCGCTCGTTCCAC 960
Qy 961 CCGTCTCTGACAGCGTCAAGGCGCTCAAGGCGCGCATGCTGAGCGCGCGCAATC 1020
Db 961 CCGTCTCTGACAGCGTCAAGGCGCTCAAGGCGCGCATGCTGAGCGCGCGCAATC 1020
Qy 1021 CGGAGCGCTCTGAGGCGAGGCGGCTTTCGCTCAACGAGAGAGAGTCAAGGTCAAG 1080
Db 1021 CGGAGCGCTCTGAGGCGAGGCGGCTTTCGCTCAACGAGAGAGAGTCAAGGTCAAG 1080
Qy 1081 GAGGAGAGAGGAGTCTTCGCGGAGAGCGGCTACCGGCTCGGAGTCAAGGTCAAG 1140
Db 1081 GAGGAGAGAGGAGTCTTCGCGGAGAGCGGCTACCGGCTCGGAGTCAAGGTCAAG 1140
Qy 1141 GCGCGGCTCTGAGGAGCATGATTCAGCGCGGCGGCTCTCTCTCTCTCTCTCTCTCT 1200
Db 1141 GCGCGGCTCTGAGGAGCATGATTCAGCGCGGCGGCTCTCTCTCTCTCTCTCTCTCT 1200
Qy 1201 TCGAGAGCGAGCAACCGCTCATCGAGTCGAGAAACAAGANGACCAACGCGGCGAAC 1260
Db 1201 TCGAGAGCGAGCAACCGCTCATCGAGTCGAGAAACAAGANGACCAACGCGGCGAAC 1260
Qy 1261 TTCCAGGCGGCTCTGAGGAGCAAGAGTGAAGAGAGTCTGAGTCTGAGTCTGAGTCT 1320
Db 1261 TTCCAGGCGGCTCTGAGGAGCAAGAGTGAAGAGAGTCTGAGTCTGAGTCTGAGTCT 1320
Qy 1321 GCGAGAGTCAACTTCAAGGAGTCAAGGAGTCTCAAGCGCGGAGTCAAGCGCGGCTCT 1380
Db 1321 GCGAGAGTCAACTTCAAGGAGTCAAGGAGTCTCAAGCGCGGAGTCAAGCGCGGCTCT 1380

Qy 1381 CCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Db 1381 CCGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
Qy 1441 GCGGCGGCGGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 1500
Db 1441 GCGGCGGCGGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 1500
Qy 1501 CAGCGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 1501 CAGCGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Qy 1561 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Db 1561 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Qy 1621 CAGCGGCTGAGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
Db 1621 CAGCGGCTGAGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
Qy 1681 GCGGCGGCTGAGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 1681 GCGGCGGCTGAGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Qy 1741 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
Db 1741 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
Qy 1801 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
Db 1801 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
Qy 1861 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Db 1861 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Qy 1921 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 1921 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Qy 1981 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
Db 1981 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
Qy 2041 TCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
Db 2041 TCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
Qy 2101 TCGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
Db 2101 TCGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
Qy 2161 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Db 2161 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Qy 2221 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Db 2221 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Qy 2281 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
Db 2281 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
Qy 2341 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400
Db 2341 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400
Qy 2401 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460
Db 2401 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460

RESULT 2
 US-09-624-693A-18
 Sequence 18, Application US/09624693A
 Patent No. 6353468
 GENERAL INFORMATION:
 APPLICANT: Yoshida, Roberta
 APPLICANT: Kocotera, Anna
 TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
 TITLE OF INVENTION: Using Same
 FILE REFERENCE: 29479/508NSC
 CURRENT APPLICATION NUMBER: US/09/624,693A
 NUMBER OF SEQ ID NOS: 25
 SOFTWARE: Patent Ver. 2.0
 SEQ ID NO 18
 LENGTH: 2439
 TYPE: DNA
 ORGANISM: Rhodotomula toruloides
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(2148)
 US-09-624-693A-18
 Query Match 68.7%; Score 1699.4; DB 4; Length 2439;
 Best local Similarity 79.4%; Pred. No. 3.2e-305;
 Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

1 ATGGCCCTCTCCCTGATCGATCGGACCTCGGTCGCAACGCGGNTGCAAGGNTG 60
 1 ATGGCACCTCTCGCTCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 60
 61 CAGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
 61 CAGGCTGT-----CAATGGGCGCTCGACCACTT-----GAGTCGAG 100
 121 GCTGCTCTCTCCGACCGACCGACCGACCGACCGACCGACCGACCGACCGACCG 180
 101 GCTGCGACCTGCGCGACCGACCGACCGACCGACCGACCGACCGACCGACCG 160
 181 ACCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 161 CGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 217
 241 TCAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 218 TCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
 301 AAGATCGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 277 AAGATCGACCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 336
 361 ACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
 337 ACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
 421 GCGTCTCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 397 GCTCTCTCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456
 481 GAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
 457 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 516
 541 GTCAGCTGCTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 517 GTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
 601 AACTTCGCGACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 577 AACTTCGCGACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636

651 GCGGACCTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 637 GCGGACCT 696
 721 GTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
 697 GTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756
 781 TTNGATCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 757 TTCAACCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 816
 841 GCGGCT 900
 817 GCGGCT 876
 901 TCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 877 TCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 936
 961 CCNTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 937 CCCTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 996
 1021 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 997 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1056
 1081 GAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
 1057 GAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1116
 1141 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
 1117 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
 1201 TCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
 1177 TCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1236
 1261 TTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
 1237 TTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1296
 1321 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
 1297 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1356
 1381 CCNTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
 1357 CCCTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1416
 1441 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
 1417 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
 1501 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
 1477 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1536
 1561 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
 1537 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1596
 1621 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
 1597 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1656
 1681 CCNTTCTCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1740
 1657 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716
 1741 CTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800

Db 1051 CGCAGCTCTTCCGCGCTGCTTTCGCGTTGAGCAGCAGAGAGAGAGTCAAGTCAAG 1110
 QY 1081 GACGACGAGAGGATTTCTCCGCGAGAGACCGCTACCGGCTCGACGCTGCTAGTGGCTC 1140
 Db 1111 GACGACGAGAGGATTTCTCCGCGAGAGACCGCTACCGGCTCGACGCTGCTAGTGGCTC 1170
 QY 1141 GCGCCGCTGCTGACGCAATGATTACGCGCCACGCGTCTCTGCTGAGGCGGAG 1200
 Db 1171 GCGCCGCTGCTGAGAGCAGATGATGCAAGCTTCACTCTCTGCTGAGAGAAC---AAC 1227
 QY 1201 TCGACGACGCAACACCGGCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 Db 1228 AGCAGACGCAACACCGGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1287
 QY 1261 TTCCAGGCGGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 Db 1288 TTCCAGGCGGCGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1347
 QY 1321 GGCAGAGCTCAACTTCAGCAGAGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 Db 1348 GGCAGAGCTCAACTTCAGCAGAGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1407
 QY 1381 CCMTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
 Db 1408 CCMTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
 QY 1441 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 1468 CACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
 QY 1501 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
 Db 1528 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1587
 QY 1561 ACGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 1588 ACTGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
 QY 1621 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 Db 1648 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1707
 QY 1681 CCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 Db 1708 CCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1755
 QY 1741 CTCGAGGCAAGAGTCAACAAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 Db 1766 CTCGAGGCTGAGGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1815
 QY 1801 CTCGAGGCTGAGGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 Db 1816 CTCGAGGCTGAGGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875
 QY 1861 NGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
 Db 1876 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1933
 QY 1921 CCGCGGAG 1980
 Db 1944 CCGCGGAG 1993
 QY 1981 CCGCGGAG 2040
 Db 1994 CCGCGGAG 2053
 QY 2041 TCGCGGAG 2100
 Db 2054 TCGCGGAG 2113
 QY 2101 TCGCGGAG 2160
 Db 2114 AGAGGATGCGGAG 2173

QY 2161 AGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 Db 2174 AGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2232
 QY 2221 NNN 2280
 Db 2233 ACCCGAG 2292
 QY 2281 TNN 2340
 Db 2293 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2352
 QY 2341 NNN 2400
 Db 2353 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2411
 QY 2401 NNN 2411
 Db 2412 AAAAAAAAAA 2419

RESULT 4
 US-09-627-216A-9
 ; Sequence 9, Application US/09627216A
 ; Patent No. 6368837
 ; GENERAL INFORMATION:
 ; APPLICANT: Sartisiani, Sima F
 ; APPLICANT: Tang, Xiao-Song
 ; APPLICANT: Qi, Wei Wei
 ; APPLICANT: Vannelli, Todd
 ; APPLICANT: Gatenby, Anthony
 ; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
 ; FILE REFERENCE: BC1009 US NA
 ; CURRENT APPLICATION NUMBER: US/09/627,216A
 ; PRIOR FILING DATE: 2000-07-27
 ; PRIOR FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 9
 ; LENGTH: 2151
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: mutant from
 ; OTHER INFORMATION: Rhodotorula glutinis
 US-09-627-216A-9

Query Match 65.7%; Score 1626.4; DB 4; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 9.3e-292;
 Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;
 QY 1 ATGGCCCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 Db 1 ATGGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
 QY 61 CAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
 Db 61 CAGCGCGT-----CATGGGCTCTGAGCAAACTC-----GCAGTGGCAG 100
 QY 121 GCTGCTCTCTCCGACCAACCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 Db 101 GCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 160
 QY 181 ACCGACCGGAG 240
 Db 161 CCGCGAG 217
 QY 241 TCGCGCGGCGGAG 300
 Db 218 TCTGCGCGGAG 276

301 AAGATGACANAGNGTGAATTCTCCGANNNGAGCTCNAACAAGNGTCTACGNGTC 350
277 AAGATTGACAAATCGGTGAAATCTTTGGGCTCGAAATCTTCATGAGGCTTACGAGGTC 336
361 ACGACTGTTTGGGCGGCTCGGCGCAACCCGGAATGAGATGCAATCTCGCTCCAGAG 420
337 ACGACTGAGTTTGGCGGATCCGACAGACCCGCAACCGAGAGCGCATCTCGCTCAGAG 396
421 GACCTCTGAGAGACAGCTGTGCGGTGTCCTCCNAAGTCATGANTCTCTGAGCTC 480
397 GCTCTCTGAGACACAGCTGTGCGGTGTCCTCTGCTGCTGAGATCTGCTCGGCTC 456
481 GGNCGGCGCTCGAGAACTGCTTCGCTCGAGTCTCGGCGGCGCATACCATTCGCG 540
457 GCGCGGCTCTCGAGAACTGCTTCGCTCGAGTCTCGGCGGCGCATACCATTCGCG 516
541 GTTAACTCGCTCAGCGCGGCGCACTCGGNGTCCGCTGCTGCTCGAGGCGCTGACC 600
517 GTTAACTCGCTCAGCGCGGCGCACTCGGNGTCCGCTGCTGCTGCTGCTGCTGCTG 576
601 AACTCTCTCAACCAACGAGCAATCAACCCCATGTCCTCTCGGCGCAATCTCGGCTG 660
577 AACTCTCTCAACCAACGAGCAATCAACCCCATGTCCTCTCGGCGCAATCTCGGCTG 636
661 GGGGACTCTCTCCCT 720
637 GGGGACT 696
721 GTTAACT 780
697 GTTAACT 756
781 TTNGTCTGAGGCGGCT 840
757 TTCAACTCTGAGGCGGCT 816
841 GCGCTCTCGGCT 900
817 GCGCTCTCGGCT 876
901 TCGGAGGCGCTCAACGCT 960
877 TCGGAGGCGCTCAACGCT 936
961 CCNTTCTCTCAACGCT 1020
937 CCNTTCTCTCAACGCT 996
1021 CGGAGGCT 1080
997 CGGAGGCT 1056
1081 GAGGAGGCGGCT 1140
1057 GAGGAGGCGGCT 1116
1141 GAGGAGGCGGCT 1200
1117 GAGGAGGCGGCT 1176
1201 TCGAGGAGGCGGCT 1260
1177 TCGAGGAGGCGGCT 1236
1261 TCGAGGAGGCGGCT 1320
1237 TCGAGGAGGCGGCT 1296
1321 GCGAGGCT 1380
1297 GCGAGGCT 1356
1381 CCNTCT 1440

1357 CCT 1416
1441 GCGGCGGCGGCT 1500
1417 GCGGCGGCGGCT 1476
1501 GCGGCGGCGGCT 1560
1477 GCGGCGGCGGCT 1536
1561 GCGGCGGCGGCT 1620
1537 GCGGCGGCGGCT 1596
1621 GCGGCGGCGGCT 1680
1597 GCGGCGGCGGCT 1656
1681 GCGGCGGCGGCT 1740
1657 GCGGCGGCGGCT 1716
1741 GCGGCGGCGGCT 1800
1717 GCGGCGGCGGCT 1776
1801 GCGGCGGCGGCT 1860
1777 GCGGCGGCGGCT 1836
1861 GCGGCGGCGGCT 1920
1837 GCGGCGGCGGCT 1888
1921 GCGGCGGCGGCT 1980
1886 GCGGCGGCGGCT 1945
1946 GCGGCGGCGGCT 2005
2041 GCGGCGGCGGCT 2100
2006 GCGGCGGCGGCT 2065
2101 GCGGCGGCGGCT 2160
2066 GCGGCGGCGGCT 2125
2161 GCGGCGGCGGCT 2186
2126 GCGGCGGCGGCT 2151

RESULT 5
US-09-765-873A-9
Sequence 9, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: B01009 US CIP
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 9

LENGTH: 2151
 TYPE: DNA
 ORGANISM: mutant from Rhodotorula glutinis
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(2151)
 US-09-765-873A-9

Query Match 65.7%; Score 1626.4; DB 4; Length 2151;
 Best Local Similarity 83.7%; Pred. No. 9.3e-292;
 Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGGCCCCCTCCCTCGACCTGATCGGACCTGAGTCCGCAAGGCTGCTGCAAGGTTG 60
 DB 1 ATGGCACCTCGCTCGACCTGATCTGCACTGTTGCAAAAGGCGTGCATCCGCAAG 60
 QY 61 CAGCCGANNCCGNNNNNGCNCAGCGGCGCCAGCTCCACCTGCTGNCGCGCCGCG 120
 DB 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GCAGTGCGAG 100
 QY 121 GCTTCCTCTCCCGACCAACCGANNAGACGACCTGACATGCTGAGNAGATCTTCGCG 180
 DB 101 GCTGCACTGCGCCACDAACCCAGGTCAAGAGTGCATGCTGAGAAATGCTCGCG 160
 QY 181 ACCCCACCGNNACGNAAGNNNTGAACTGAGCGGTACACCTCTCACCTCGGAGCTG 240
 DB 161 CGCGACCTGACTGA---CGCTGAATCTGACGGCTACCTGCTCAACTCGGAGAGCTG 217
 QY 241 TCGCGCGCGCGCGAGGCGCGCNGTCCGCTGNCAGACAGNAGAGATCCGCGCA 300
 DB 218 TCTCGCGCGGAGAGAGGCGAGGCTGCTCGCTCAAGAGACG--CGAGAGATCCGCTCA 276
 QY 301 AAGATCGACAAANAGTTCGAGTTCCTCCGNNNCAGCTCNAACAAGNGTCTACGAGT 360
 DB 277 AAGATTACAAATCGGTGAGTTCTTGCGCTGCAACTCTCATAGAGGTCTACGCGCTC 336
 QY 361 ACAGACTGTTTCGGCGCTCGGCGACACCCGAGCTGAGAGATGCAATCTGCTCCAGAG 420
 DB 337 ACAGCTGAGATTTGCGGATTCGCGACACCCGACCGAGAGAGGCAATCTGCTCCAGAG 396
 QY 421 GCNCTCTCGAGACAGCTCTCGGCTGCTCCGNNNCAGCTCNAACAAGNGTCTTNGCTC 480
 DB 397 GCTTCCTCGAGACAGAGCTCTGCGGTCTTCCTCGCTGCTGCTGCTGCTGCTGCTGCT 456
 QY 481 GAGCGCGCTCGAGAACTGCTGCTGCTGAGAGTTCGCGCGCGCATACATCCGCG 540
 DB 457 GAGCGCGCTTCGAGAACTGCTTCCCTCGAGAGTTCGCGCGCGCATACATCCGCG 516
 QY 541 GTCAACTGCTCAACGCGCGGCACTCGGCGGTTCGCGCTGCTGCTGAGAGCTGCAAC 600
 DB 517 GTCAAGCTGAGCCGCGCGCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 576
 QY 601 AACTTCTCAACACGAGCATACCCCATGCTGCTGCTGCGGCGCATCTCGGCTG 660
 DB 577 AACTTCTCAACACGAGCATACCCCATGCTGCTGCTGCGGCGCATCTCGGCTG 636
 QY 661 GGAGACCTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 DB 637 GGAGACCT 696
 QY 721 GTNCACTNNNTNACGAGGCGANNAGAAATCATGTGNCGCCGCGAGGCGATCGCGCT 780
 DB 697 GTGACAGCTGCTCAAGAGGCGAGAGAAATCTGTACGCCCGAGAGGCGATGCGCTC 756
 QY 781 TTNAGTTCGAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 DB 757 TTCAACTCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 816
 QY 841 GCGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 DB 817 GCGCTCTCAAGCTGAGAGGCGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876
 QY 901 TCGAGAGGCGCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

DB 877 TCGAGTGTGCTCAAGGCGCATGAGGTGAAAGCGATGTGCGCAACCGCGCTGCTTCCAC 936
 QY 961 CMTTCCGCGACGAGCTGACGCGCGCTGACCGGACGATGAGGTGCGCGCAACATC 1020
 DB 937 CCTTCTTCAAGAGCTGACGCGCGCTGACCGGACGATGAGGTGCGCGGAAACATC 996
 QY 1021 CGACGCTCTCGAGGCGAGCANNATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 DB 997 CGAAGCTCTCGAGGAGAACCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1056
 QY 1081 GACGACGAGGCGATTTCTCGCGACGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 DB 1057 GACGACGAGGCGATTTCTCGCGACGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1116
 QY 1141 GCGCGCTGCTGACGAGCATGATTTACCGCGACGCGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 DB 1117 GCGCGCTGCTGACGAGCATGATTTACCGCGACGCGCTGCTGCTGCTGCTGCTGCTGCTG 1176
 QY 1201 TCGAGACCGGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 DB 1177 TCGAGACCGGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1236
 QY 1261 TTCAAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 DB 1237 TTCAAGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1296
 QY 1321 GCGAGGCTCACTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 DB 1297 GCGAGGCTCACTTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1356
 QY 1381 CMTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
 DB 1357 CMTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1416
 QY 1441 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 DB 1417 GCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1476
 QY 1501 CAGCGCGGAGATGAGGCAACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
 DB 1477 CAGCGCGGAGATGAGGCAACAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1536
 QY 1561 ACNGCGGAGGCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
 DB 1537 ACAGCGAGTCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1596
 QY 1621 CAGGCGCTGCACTCGCGGAGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
 DB 1597 CAGGCGATCGACTTGGCGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1656
 QY 1681 NCNNGCTGCTGAGGAGCACTTTGCGCNGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
 DB 1657 GTCTGCTGATGAGCAAGCACTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1716
 QY 1741 CTCGNGACAAGTCAACAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
 DB 1717 CTCGNGACAAGTCAACAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1776
 QY 1801 CTGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 DB 1777 CTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1836
 QY 1861 NGTCTGCTGCGCANNAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
 DB 1837 TCGTGAAGT-----GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1885
 QY 1921 CGCGGAGAGGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 DB 1886 CGCGGAGTGGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1945
 QY 1981 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040

1081 GACGACGAGGAGGATTTCTCCGACGAGACGCTACCCGCTCCGCAAGTGCCTCAGTGGCTC 1140
Db 1057 GACGACGAGGAGGATTTCTCCGACGAGACGCTACCCCTTGCCGACGCTCCTCAGTGGCTC 1116
Qy 1141 GGGCCGCTCGTCAAGCAGATGATTAACGCGCCACGCGNCTCTCTGCTCGAGGCGGACG 1200
Db 1117 GGGCCGCTCGTCAAGCAGATTAACGCGCCACGCGCTCTCAGCAGTGAAGGCGGCGAG 1176
Qy 1201 TCGACGACGAGACACCCGCTCATTCAGCTCGAGAACAGAGACCCCAACGAGGCGGAC 1260
Db 1177 TCGACGACGAGACACCCCTCTCATTCAGCTCGAGAACAGAGCTTGACACCGGCGGAA 1236
Qy 1261 TTCGAGGACGACGCTCTCGACCAACGATGAGAGAGTGGCTCGCTCGCTGCTGATC 1320
Db 1237 TTCGAGGACGACGCTCTGCGCAACGATGAGAGAGTGGCTCGCTGCTGCTGCTGATC 1296
Qy 1321 GGGCAGCTCACTTCAAGCAGCTCAACCGAGATGCTCAACCGCGGATGAGAGCGGCGCT 1380
Db 1297 GGGCAGCTCACTTCAAGCAGCTCAACCGAGATGCTCAACCGCGGATGAGAGCGGCGCT 1356
Qy 1381 CCNTCCGCTCGCTCGGACGAGACCCGCTCTCTCTATCACTGCAAGGCGCTGAGATT 1440
Db 1357 CCTCTGCTCGCTCGGACGAGACCCGCTCTCTCTACCACTGCAAGGCGCTGAGATT 1416
Qy 1441 GCGCGACGACGCTCACTTGGAGCTCGAGACCTTGGCAACCGGCTTACGACCAAGCTC 1500
Db 1417 GCGCGCTCGGCTGACCACTCTCGAGTGGAGACCTTGGCAACCGGCTTACGACCAAGCTC 1476
Qy 1501 CAGCGCGACGATGAGGACCAACGAGCGCTCACTGCTCGGCTGATCTCGGCGGCGC 1560
Db 1477 CAGCGCGCTGAGATGAGGACCAACGAGCGCTCACTGCTCGGCTGATCTCGGCGGCGC 1536
Qy 1561 ACNCGGACGACCAAGCAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Db 1537 ACNCGGACGACCAAGCAGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
Qy 1621 CAGCGCGTCACTCTCGGCGGATGAGGCTGAGGCTCAAGAGCAGTTCGACCGGCTNN 1680
Db 1597 CAGCGCTCACTTGGCGGATGAGGCTGAGGCTCAAGAGCAGTTCGACCGGCTNN 1656
Qy 1681 NCNCGCTCTCTNAGGAGCACTTGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1740
Db 1657 GTCGCTCATTCAGACGAGCACTTGGCTCTCGGCTCAAGCAGGCTCGACCTGCGAG 1716
Qy 1741 CTGCGAGCAAGGCTCAAGAGGCTCTCAAGGCTCTGAGCAGACCACTCGGAGAC 1800
Db 1717 CTGCTGAGAAAGGTGAACAGAGCTGCGCAAGCGCTCTGAGCAGACCACTCGGAG 1776
Qy 1801 CTGAGCGGCTGAGCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1860
Db 1777 CTGCTCTCGGCTGAGCAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1836
Qy 1861 NNGTCTCTGCGGCGGCTCTGCTGCGGCTCTGAGCGGCTCTGAGCGGCTCTCTCTCT 1920
Db 1837 TCGTCTGAGCT 1885
Qy 1921 CCGCGAGAGGCGCT 1980
Db 1886 CCGCGAGAGGCGCT 1945
Qy 1981 CCGTCTGCT 2040
Db 1946 CCGTCTGCT 2005
Qy 2041 TCGCGAGAGGCT 2100
Db 2006 TCGCGAGAGGCT 2065
Qy 2101 TGAAGATCGGACCAAGCT 2160
Db 2066 TGAAGATCGGCT 2125

Qy 2161 ACGTCTCTCTCAAGATGCTCGCTAG 2186
Db 2126 ACGTCTCTCTCAAGATGCTCGCTAG 2151
RESULT 8
US-09-765-873A-31
Sequence 31, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BC1009 US CIP
CURRENT APPLICATION NUMBER: US/09/765,873A
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 31
LENGTH: 2061
TYPE: DNA
ORGANISM: mutant from Rhodotorula glutinis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2061)
US-09-765-873A-31
Query Match 65.1%; Score 1610.4; DB 4; Length 2061;
Best Local Similarity 85.0%; Pred. No. 8.2e-289;
Matches 1764; Conservative 0; Mismatches 297; Indels 15; Gaps 3;
Qy 111 GCCGCGGAGGCTGCT 170
Db 1 GCAGTGGAGGCTGCT 60
Qy 171 ATCTCTGCGGACCCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 230
Db 61 ATCTCTGCGGACCCGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 117
Qy 231 GGNAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 230
Db 118 GGNAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 176
Qy 291 GATCGCGGAGGATGCAAGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 350
Db 177 GATCGCGTCAAGATGCAAGAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 236
Qy 351 CTACGAGGCTCAGACTGCTTTCGCGGCTGCGGCGGACCGGACCGGACCGGACCGGACCG 410
Db 237 CTACGAGGCTCAGACTGCTTTCGCGGCTGCGGCGGACCGGACCGGACCGGACCGGACCG 296
Qy 411 GCTTCAGAGGCT 470
Db 297 GCTTCAGAGGCT 356
Qy 471 CTTGCGCTCTGAGCGGCTCTGAGACTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 530
Db 357 GTTCGCGCTCTGAGCGGCTCTGAGACTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 416
Qy 531 GACCATCTGCTCACTGCTCTCAAGCGGCGGACCGGACCGGACCGGACCGGACCGGACCG 590
Db 417 GACCATCTGCTCACTGCTCTCAAGCGGCGGACCGGACCGGACCGGACCGGACCGGACCG 476
Qy 591 GGGCTCAGCACTTCTCTCAAGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 650
Db 477 GGGCTCAGCACTTCTCTCAAGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 536
Qy 651 CTGCGGCTGAGGCT 710
Db 537 CTGCGGCTGAGGCT 596

```
QY 711 GGAACNNCAAGTNCACGTTNTNCAAGAGGSCANNNGAAGATCATGTNCGCCCGGAGGC 770
Db 597 GGAACAGAAAGTGCAGTGCCTCCACAGAGGSCAAGAGAAAGTCTGTAGCGCCCGGAGGSC 656
QY 771 GATCGGCTCTTNGTCTCGAGGCCGTCTCTCGGCCCGGAAGAGAGGTCTCGTCTCGT 830
Db 657 GATGGCGCTCTTCAACCTCGAGCCGTCTCGGCCCGGAAGAGAGGTCTCGTCTCGT 716
QY 831 CAACGGACAGGCGGTCTCGGCGCTCGATGGCGAACCTCGCTCTGCGACGACGACATATCT 890
Db 717 CAAGGGACCGCGGTCTCGAGATCATGATGGCACCTCGCTCTGCGACGACGCTCATATCT 776
QY 891 CTGCTCTCTCTCGAGGCGCTCGACGCTCTMAAGGTGAGGCGCATGTCTGGCGACGCGCG 950
Db 777 CTGCTCTCTCTCGAGATCGCTCGAGCGCATGACGCGTGAAGGATGTCTGGCGACGCGCG 836
QY 951 CTGCTCTCAACCCNTTCTCCACAGCGTCAAGGCGCTTCCACCCGACCGCAGATGAGGTGCG 1010
Db 837 CTGCTCTCAACCCNTTCTCCACAGCGTCAAGGCGCTTCCACCCGACCGCAGATGAGGTGCG 896
QY 1011 GCGCAACATCCGACGCTCTCGAGGCGCAGCANNGTTCGCGCTCCACGACGAGAGAGGT 1070
Db 897 GGGAAACATCCGACGCTCTCGAGGCGCAGCAGCTTTGCTGTCCACATGAGAGAGGT 956
QY 1071 CAAGGTCAAGAGACGAGAGGCGATTTCTCGCGAGAGCGGCTAACCGGCTCGGACGCTGCG 1130
Db 957 CAAGGTCAAGAGACGAGAGGCGATTTCTCGCGAGAGCGGCTAACCGGCTCGGACGCTGCG 1016
QY 1131 TCAGTGGCTCGGCGCGCTCGTCAAGACATGATGACGCGCCGACGCGGCTCTCGTCTCGA 1180
Db 1017 TCAGTGGCTCGGCGCGCTCGTCAAGACATGATGACGCGCCGACGCGGCTCTCGTCTCGA 1076
QY 1191 GCGCGGACGATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1250
Db 1077 GCGCGGACGATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1136
QY 1251 GCGCGGACGATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1310
Db 1137 GCGCGGACGATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1196
QY 1311 GCGCGGATCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1370
Db 1197 GCGCGGATCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1256
QY 1371 GCGCGGCTTCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 1430
Db 1257 GCGCGGCTTCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 1316
QY 1431 CCTCGACATTTGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1490
Db 1317 CCTCGACATTTGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1376
QY 1491 GACCCACGTCGACGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1550
Db 1377 GACCCACGTCGACGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1436
QY 1551 GCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1610
Db 1437 GCGCGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1496
QY 1611 CTGCGGCTCTCGAGCGGCTCGACCTCGCGCGATGAGTTCGAGTTCAGAGAGAGTTCGAG 1670
Db 1497 CTGCGGCTCTCGAGCGGCTCGACCTCGCGCGATGAGTTCGAGTTCAGAGAGAGTTCGAG 1556
QY 1671 CCGGATTTTCCGAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1730
Db 1557 CCGGAGCAATCGTCTCGCTCATGACACACACTTTGCTCTCGCCGACGACGCTCGAAGCT 1616
QY 1731 NNNNNACGACTCGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1790
Db 1617 GCGGACGAGCTCTCTGAGAGAGTGAACAAGCGCTCGCAAGCGCTTCGAGCAGACCA 1676
```

```
QY 1791 CTGTAAGACCTCGAGCGCGGCTGGACGACGACGCTCTCTCTCTCTCTCTCTCTCTCTCT 1850
Db 1677 CTGTAAGACCTCGAGCTCGCGGCTGGACGACGACGCTCTCTCTCTCTCTCTCTCTCTCTCT 1736
QY 1851 CGAGNNNNNNNNNGTCTCTCGCCNNNNNCCANNAGAGTCTCGCTCGCAGCCGTCGACGCTCG 1910
Db 1737 CGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1785
QY 1911 AAGTGGCTCTCGGCGAGAGGCGATCTCGCTCAAGCGCGCANNAGTCCGACGACGCTCTCG 1970
Db 1786 AAGTGGCGCGCGCGAGTGGCGCATCTCGCTCACCCGCGCAAGTCCGGAACCTTCGCG 1845
QY 1971 NCGGCNCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2030
Db 1846 TCGCGCGGCTGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1905
QY 2031 TACTGCTTGTCTCGGAGAGGCTCGGCTCAAGGCGCGCGCGGCGAGCTTCTCTGCGC 2090
Db 1906 TACGCTTGTCTCGGAGAGGCTTGGCGTCAAGGCGCGCGCGGAGAGCTTCTCTGCGC 1965
QY 2091 AAGCAGAGGTGACGATCGGACGACGACGCTCGGACGCTCTGACGAGGCGCATGAGNNCGGC 2150
Db 1966 AAGCAGAGGTGACGATCGGCTCGACGCTCTTCCAAAGTCTTCAAGGCGCATGAGTGGGCG 2025
QY 2151 NCATCAACGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2186
Db 2026 AGGATCAACGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2061

RESULT 9
US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koestera, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Rhodotorula mucilaginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (646)..(2784)
US-09-624-693A-16

Query Match 64.6%; Score 1598.6; DB 4; Length 2787;
Best Local Similarity 82.9%; Pred. No. 1.3e-286;
Matches 1812; Conservative 0; Mismatches 330; Indels 44; Gaps 5;
```


241 TCGGCGCCGCGCAGAGGCGCGCNGTCGCGCTGACAGAGAGAGAGAGAGAGATCCGCGCA 300
881 TCGGCGCTGTGCTGCGCGCTGCGCTGCGTCAAGGTGAGAGAGCGCCGCG -ACATCCGCGAG 939
301 AAGATGACAAAGAGGTGAGTTCTCTCGGAGCAGCTGACAGAGAGAGTCTAGAGAGTGC 360
940 AAGATGAGAGAGAGGTGAGTTCTCTCGGAGCAGCTGAGAGAGAGTCTAGAGAGTGC 999
361 ACGAGTGTGTTGCGGAGCTCGGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
1000 ACGAGTGTGTTGCGGAGCTCGGCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
421 GCGCTGCTGAGAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 480
1060 GCGCTGCTGAGAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1119
481 GCGGCGGCGCTGAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 540
1120 GGTGCGGCGCTGAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1179
541 GTCAAGTCTGCTGAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 600
1180 GTCAAGTCTGCTGAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1239
601 AACTTCTCTCAACAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 660
1240 AACTTCTCTCAACAGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1299
661 GCGGAGCTCTGCGGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
1300 GCGGAGCTCTGCGGAGAGAGTCTGCGGTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359
721 GTNCAAGTCTGAG 780
1360 GTNCAAGTCTGAG 1407
781 TTNGAGTCTGAG 840
1408 AAGAGTCTGAG 1467
841 GCGGAGTCTGAG 900
1468 GCGGAGTCTGAG 1527
901 TCGAG 960
1528 GCGAG 1587
961 CCNTTCTCTGAG 1020
1588 CCNTTCTCTGAG 1647
1021 CGCAGAGTCTGAG 1080
1648 CGCAGAGTCTGAG 1707
1081 GAGCAG 1140
1708 GAGCAG 1767
1141 GCGCGGCTGCTGAG 1200
1768 GGTGCGGCTGCTGAG 1827
1201 TCGAG 1260
1828 TCGAG 1887
1261 TTCCAG 1320
1888 TTCCAG 1947
1321 GGCAAGTCTCAAGTCTCAAG 1380

1948 GGCAAGTCTCAAGTCTCAAG 2007
1381 CCNTTCTCTGAG 1440
2008 CCNTTCTCTGAG 2067
1441 GCGGAG 1500
2068 GCGGAG 2127
1501 CAGGCGGAG 1560
2128 CAGGCGGAG 2187
1561 ACNCGAG 1620
2188 ACNCGAG 2247
1621 CAGGCGGAG 1680
2248 CAGGCGGAG 2307
1681 NCNCGAG 1740
2308 ACTGAG 2352
1741 CTGAG 1800
2353 GTGAG 2412
1801 CTGAG 1860
2413 CTGAG 2471
1861 NNCGAG 1920
2472 -----CGCGCGAG 2521
1921 CGCGCGAG 1980
2522 CGCGCGAG 2581
1981 CGT 2640
2582 CGT 2641
2041 TCGGAG 2100
2642 TCGGAG 2701
2101 TGACAG 2160
2702 TGACAG 2761
2161 AGCTGT 2186
2762 CCGTCTCTGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2787

RESULT 10
US-09-624-693A-14
Sequence 14, Application US/09624693A
Patent No. 635468
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide sequences and Methods of Obtaining and
FILE REFERENCE: 29479/50NSC
CURRENT APPLICATION NUMBER: US/09/624,693A
CURRENT FILING DATE: 2000-07-24


```

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2311
; TYPE: DNA
; ORGANISM: Amanita muscaria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)..(2237)
US-09-624-693a-14

Query Match      13.0%; Score 322.6; DB 4; Length 2311;
Best Local Similarity 52.8%; Pred. No. 3e-51;
Matches 734; Conservative 0; Mismatches 625; Indels 30; Gaps 4;

QY 345 CAGNGCTTACGAGTACGACTGTGTTGCGCGCTTCGCGCAGACCCGAGTGAAGATGC 404
DB 368 CACGATATACGAGCTGTCAACTGTGTTGCGTGGCAGTGTGATACAGCGACGACAAACC 427
QY 405 NATCTGCTCCAAAGAGCNCCTCTCGAGACAGCTGTGCGGTGTCCTCCAGCTGANT 464
DB 428 GATGTTTGTGGGTTTGCCCTTTTGCAACACAACTGTAGGATATCTGCCACCTCGAC 487
QY 465 CGANTCCTTCNGCCTCGAGCGCGGCGCTCGAG-----AACTGCTTCGCTCGAGGT 515
DB 488 TGAACCTTTGAGCGTCTACTCTCTCAAGATGCAATAACAGACATGCCAGAGGCGTG 547
QY 516 CGTCCGGGGGCGCATACCACTCCGCGTCACTCGCTCANCGCGGCGCACTGGCNGTCCG 575
DB 548 GATTCGGGGGGCATTTTATCCGATATGAAATTCGTAATTCCTGGCCACTCTGGAAATAG 607
QY 576 CCTGCTGCTCTCGAGCGCTCCCAACTTCTCTCAACAGGAGATCAACCCCATGCTCC 635
DB 608 ATGGAGTTGATGAAAAGATGAGAACTACTCGGGCGCAATGTGATACCTGTCTCC 667
QY 636 CTTCCGCGGACCATCTGGCGCTCGGCGCACTCTCCCGCTCTCTAATATCGCGCGCC 695
DB 668 CCGAAGAGGAGCATCTCTCACTCGGAGATCTGTCTCCCTATCTATATCGAGGAC 727
QY 696 CATCACCGGTCAACCCGAGCANNAGGTCAGCTNNTC-----ACGAGGAGANNAGAA 749
DB 728 GATTATTTGGCAACCCATCAATCAAGTATATACAGCTCATTAAGTCCGAAATTCGCCA 787
QY 750 GATCATGTGCGCGCGGAGGAGATCGCGCTCTTNGGTCTCGAGCCCGTCTGCTCGGCC 809
DB 788 AATGTGATCCTCGAAGATGTCTTGCTGTGATATATCGAACTTCCACATCGAATC 847
QY 810 GAAGAGGGGTCTCGCTCTGTCTCAACGCGACCGGCGCTTCGCTGTGATGGGACCCCTGCC 869
DB 848 GAAAGAACTCTTGATTTTAAATGGACCGCATCTCGGATCTGGGAGCTTTAGC 907
QY 870 TCTGACAGACGACATCTCTCGCTCTCTCGAGGCGCTCAAGGCTCTTACGCTCGA 929
DB 908 CTTAAACGAAAGCATCATCTGTGTGTGTGCTGAAGTGTGACAGGCTATAGGGACCGA 967
QY 930 GGGCATGTGTGCGCGACGCGGCTGTTTCCACCCNTTCTCTCAAGAGTACGCGCCCTCA 989
DB 968 GGGATGTAGTAGGACTCGGCTTCTCATAGCACCGTTCATTTATGCGACCGACACACAA 1027
QY 990 CCCGACCCAGATCGAGGTGCGCGGCAACATCGCGACGCTCCCTCGAGGGCAGCANNATTGC 1049
DB 1028 TCCCGGTCAAGTAGATGTGTGAGAACTTTGGAATTTGCTCGATGGGATTAATTGGC 1087
QY 1050 CGTTCACCAAGAGAGAGAGTCAAGGTCAAGACAGACAGAGGCAATTCCTCGCCACGACG 1109
DB 1088 TCAGTTAGAAAGACAGAGTTGCTGCTAGAAAGACGATTAATACCCCTTCGCGACGACG 1147
QY 1110 CTACCGCGCTCCGAGGTGCGCTCAAGGTCTGCGCGCGCTCTCTGACGACATGATTCACGC 1169
DB 1148 TTTTCCACTCCGAACTTCCCTCTAATCTTGTGGCTTAAGATGAACATTAATCTCCGC 1207
QY 1170 CCAAGCNGTCTCTGCTCTGAGGCGGACAGT---CGACGACCGACCAACCGCTCATGCA 1226

```

```

DB 1208 TTTCCAGACTGTAAACGAGAGTGTAAATTACTTAACAGTACTGCAAAATCCATGTGA 1267
QY 1227 CGTCGAGAACAAAGANAGACCCACACGCGCGCAACTTCGAGCGAGCGGCTGGCNAACAC 1286
DB 1268 TGGGAGACTGCGGGAATCTCACACGCTGTGCAATTTCCAAAGGATGGCTGTACTAATCC 1327
QY 1287 GATGAGAAAGACTCGCTCTGCGCTGTGCGCTGTGATGGGCAAGCTCAACTTCAGCAGCTAC 1346
DB 1328 AATGAGAAAGACGCGACTTGTCTTACATACGTTGSCAAATTAATTTCCAGAGCAC 1387
QY 1347 CGAGATGCTCAACGCGGCGATGACCGCGGCGCTGNCNTCTGCGCTGTGCGAGGAGCC 1406
DB 1388 TGATTTAGTCAATCTGCGATTAACCGCGCTTTCGCGCTTCACTGAGTGCACAGATCC 1447
QY 1407 NTGCTCTCTTATCACTGTGCAAGGCGCTGACATTTGCGCGNGCGNTACATTCGAGCT 1466
DB 1448 ATCTCACTACAGCGCCAAAGACTAGACATAGCACTGGGCGCTTACGAGCGA--- 1504
QY 1467 CGGNAACCTTGGCAACCGGTTAAGACCCAGCTCCAGCGCGGCGNGAGATGGGCAACAGGC 1526
DB 1505 -----AGGACTCTGCGCCCACTCACTGATGAGCGAAGAAATGCAACCAAGC 1555
QY 1527 CGTCAACTGCTCGCNCCTCATCTCGGCGCGCGCACGCGGCAAGGCAACGACTCTTTC 1586
DB 1556 TGTTAATCTCTGGGCTGATTTTGTCTCGGCTACATCAATGCTTGAAGTCTAAC 1615
QY 1587 TCTCTCTCTGCGCACCACTCTACTGCGTCTCGAGCGCTCGAAGCTTCGCGGATGGA 1646
DB 1616 ATCTGATGAGCGGTCTTACTGTATATTTCTATGCGCAACTCTGACCTCCGCTGCTTCA 1675
QY 1647 GTTCGAGTTCAAGAAAGAGTTCGACCCGNTNNTGNCNNGCTCMTCAAGAGCACTTTGG 1706
DB 1676 GCGGAGTTCTTGCGCGGCTGTAGACATCATCTGTGTAGAGATTAAGATGCTATTGG 1735
QY 1707 CNGCGCCT 1715
DB 1736 ATCTTTCCT 1744

RESULT 11
US-09-615-192A-245
; Sequence 245, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 06/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 06/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 245
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-615-192A-245

Query Match      8.6%; Score 213.2; DB 4; Length 1455;
Best Local Similarity 54.8%; Pred. No. 4.4e-31;
Matches 480; Conservative 0; Mismatches 384; Indels 12; Gaps 4;

QY 465 CGANTCCTTCNGCCTCGAGCGCGCTCGAGACTTCCTCGCTCGAGTGTGTCGCGG 524
DB 592 CCGATCTTCGCAACGAGCAGAGTGTGTCACACACCTGCTCAATCTCCACCGAGC 651

```

QY	522	CGCCATGACCATCCGGGTAACTCCGGTACACGCGGCACATCCGCGCTCGTGC	584
Db	652	CGCCATGCTGTCCGGGTAAACACCTCTCTCCAGGGCTAATCCGCAATCGTTTGAGAT	711
QY	585	CCATGAGCGCTCACCACTTCTCTCAACCAAGGCATCAACCCCATGTCCTCCCGCG	644
Db	712	CTCTGAGGCGCATCAACCAAGTTCCTCAACCAACATCAACCCCGTGCCTCCAGGGG	771
QY	645	CACCATCTCGGCGTCCGGGGGACCTCTCCCGCTCTCAACATCCGCGCGCATCAACGG	704
Db	772	CACCATCACTGCTTCAGGCGACTGTGTTCCCTCTCTCAATTCGCGGGCTTCAAGGG	831
QY	705	TCACCCGGAACNNCAAGTTCACGTTNNNTNACAGAGGCGANNAGAAATCATGTCGCCG	764
Db	832	CCGGCCCAATCCAAAG---CCGTGGGGCTGATGGGAAGTCCCTGACGCTGTCGAGGC	888
QY	765	CGAGGCGATGCGCTCTTNGTGTGAGCCCTGCTCTCGGCGCCGAAGAGGGTCTCGG	824
Db	889	CTTCCGGCTTCGCGGATGGAACAAGGCTTCTTTCAGCTGACGCCCCAAGAGAAAGGTTGGC	948
QY	825	TCTGTCACGCGACGCGCGCTCTCCGCTCTCAATGCGACACCTCGCTTGCACGACACA	884
Db	949	GCTGTGACAGGACGCGACCTCGGGTCTGCGCTGCTTCATCTGCTCTTCGAGGCCAA	1004
QY	885	CATGCTCTGCTCTCTCTCGAGGCGCTCAAGCTCTCAAGGTTCAGGCTCATGTGTCGCA	944
Db	1009	CATACTCGCGGCTCTCTCGAGGCTCTCTCAACGATCTTCGAGAGGTGATGACAGGGAA	1060
QY	945	CGCGGCTGTGTCACCCNMTCTCTCAAGAGTCAAGCGGCTTCAACCGACCCAGATCGA	1000
Db	1069	GCGGAGTTCAAGACCACTTGAACGATTAATTGAAGACCAATCCCGG---CAGATTGA	1124
QY	1005	GGTGCGCGCACATCCGACGCTCTCTCGAGGCGCAANNNGTTTCCGTTCACCAAGGA	1060
Db	1126	GTCTGGGGCTAATATGAGACAATTGTGATGAAAGCGC--TTCAGTGAAGGCTGCTAA	1184
QY	1065	GGAAGTCAGGTCAAGACGACGAGGAGCATTTCCGCGAGGACCGCTACCCGCTCCGAC	1124
Db	1183	AAAGTTGACAGAGATGATTCGCTCCAGAAAGCCAAAGACAGGACAGATACGCTTCAGGAC	1240
QY	1125	GTAGCCTCACTGCTCGGCGCGCTGCTGACGACATGATTCAGCGCCACGCGTCTCTC	1184
Db	1243	TTCCTCCCACTGGCTAGGGC---CAGATTGAGTATCCGAGCGGACCAAGATGAT	1240
QY	1185	GCTCGAGCGCGGACGTGACGACGACACACCCGCTCATCGACGTTCGAGAACAGANAGC	1240
Db	1300	TGAGAGGAATCATATTCGTTCATATACAAACCGCTGATCGATGTTCGGAGAAACAAGGC	1340
QY	1245	CCACCAAGGGGCGCACTTCAGGCGAGCGCTGTGCGAACAACGATGGAAGACTCGCT	1340
Db	1360	CTGTCAAGGTGGAACTTCAGGAGACCCCGATTGTTGCTTCATGACACACTCGCT	1400
QY	1305	CGCCTGCGCCTGATCGGCAAGCTCACTTCAGCA	1340
Db	1420	GGCGGTGCTCATATGAGGCAAGCTCATGTTGCGGA	1455
RESULT 12			
US-09-615-192A-247			
; Sequence 247, Application US/09615192A			
; Patent No. 6410718			
; GENERAL INFORMATION:			
; APPLICANT: Bloksberg, Leonard N.			
; APPLICANT: Havukala, Ilkka			
; TITLE OF INVENTION: Materials and Methods for the			
; TITLE OF INVENTION: Modification of Plant Lignin Content			
; FILE REFERENCE: 11000.100354U			
; CURRENT APPLICATION NUMBER: US/09/615,192A			
; CURRENT FILING DATE: 2000-07-12			
; PRIOR APPLICATION NUMBER: US 08/975,316			
; PRIOR FILING DATE: 1997-11-21			
; PRIOR APPLICATION NUMBER: US 08/713,000			
; PRIOR FILING DATE: 1996-09-11			

```

? PRIOR APPLICATION NUMBER: US 09/169,789
? PRIOR FILING DATE: 1998-10-09
? NUMBER OF SEQ ID NOS: 405
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO: 247
? LENGTH: 1520
? TYPE: DNA
? ORGANISM: Pinus radiata
US-09-615-192A-247

```

Query Match	6.9%;	Score 170.2;	DB 4;	Length 1520;
Best Local Similarity	49.7%;	Pred. No. 3.8e-23;		
Matches 480;	Conservative 0;	Mismatches 473;	Indels 13;	Gaps 4;

QY	791	AGCCCGCTGCTCTCGGCCCGGAAGAGGGGTCTCGCTCTCTCAACGCAACGCGCGCTCCG	853
Db	124	AGCCCTTTGAATTGAGCTTAAAGAAAGTCTGGCCATTGTCAATGGCACTTCAGTGGAG	183
QY	851	CTTCGATGCGCACTCTCGCTCTGAGCGAGCGCAACATGCTCTCGGCTCTCGCAGCGGC	910
Db	184	CAGCACTGGCTTCCATTGTGTGTTTCGATCGCAATGTTCTTGCTCTGCTCTCTGGAATGA	243
QY	911	TCACGGCTTNAAGTTCGAGCGCATGCTCGGCCACCGCGGCTGTTCACCCNTTCTCC	970
Db	244	TCCTGTCGACATGTTTCGAGAGGTATGATGGGAAGCTGATGATTAACAGATTCATTAATCTC	303
QY	971	ACGACGTACCGGGCCCTCACCCGACCCGAGATCGAGCTGGGGCGCAATCCGCACTCC	1030
Db	304	ACAAGTGAAGCACCAT---CTGGCCAAATGGAAGCTGACAGATCAATGAGATATGTCT	360
QY	1031	TCGAGGGACGAGNNGTTCGCCCTCACACGAGAGAGAGGTCAAGGTCAAAGACGACGAG	1090
Db	361	TGGACGGGA---GTCCTATATGAAACACCTCTCTAAGCTCCATGAGATGAATCCTCTGC	416
QY	1091	GCATTCCTCCGCGAGGACCGGTACCCGGCTCCGCAAGTCCGCTCAATGAGCTCCGCGCTCG	1150
Db	417	AGBAGCCAAAGCAGGATGCTATGCGCTTGCGCACTTGCTCAGTGGCTCGGCCCTCAGG	476
QY	1151	TCAGCGACATGATTACGCCGCCACGAGCTCTCTCGCTCGAGCGCGGAGCTGACGACCG	1210
Db	477	T---GGAGATTTATCAGATCTTGCAACTCACTGATGATGAGCGGGAAATCAATTGTGTAATG	533
QY	1211	ACAACCGGCTCATGACCTCGAGAAACAAGANGAACCAACGCGCGGCACTTCCAGAGCG	1270
Db	534	ACAATTCAGTAAATTAATTTATTTGCCAGAGACAAGTCTACATGAGAGGAATTTCCAGGGCA	593
QY	1271	CGGCTGTGCNPAACGATGAGAAAGACTGGCTCGGCTCGGCTCGCTCGCTGATTCGGCAAGTCA	1330
Db	594	CACCTATTGTTGTTTCCATGATATATCTTGCTGTCTCAATTTCAGCAATTTGGAAATTTGA	653
QY	1331	ACTTCAGGAGTCAACGAGATGCTCAACCGCGGCACTGAACCGCGGCTTNCNTCCCTCC	1390
Db	654	TGTTGCTCAATTTCTCABAGCTTGATGATGATTACTATCAATGAGAGCTTGCCCTCGAATC	713
QY	1391	TCGCTGCGGAG--GACCCNTCGCTCTCTCTATCTACTGCAAGGGCCCTGCACATTCGCGG	1447
Db	714	TGAGTGTGTGGCTAATCCAGCTCGATTAATGACTGAAGGGGCGCGAGATGCTGATGG	773
QY	1448	CNCGNTACACTTCGAGCTCGGNCACCTTGCCAAACCGGATNACGACCACTGTCAGACGG	1507
Db	774	CTTCTTACACTTCTGAGCTCTTTACCTGGCAATCTGTGCACCACTGATACAGAGCG	833
QY	1508	CNAGATGGGCAACGAGCGCTCAACTCGCTCGGNCATCATCTCGGNGCGCGGCAACGCG	1567
Db	834	CCGAAACGATTAACGAGATGTCATTTCTCTGGGTCTCGTTTAACTTAABAAATCTGGCG	893
QY	1568	AGGCCAAGACGCTCTTTCTCTCTCTCGGCAACCACTTAACTGCGTTCCTCAGGCG	1627
Db	894	AGGCATCATATTTCTGAAGCTGATGCTTCCATCTCTGACAGAGCTCTGTCCACAGCTG	953
QY	1628	TCGACCTCGGCGGAGTGAAGTTGAGTTCAAGAAAGAGTTCGACCGGANTNNTCNNGCG	1687
Db	954	TGGATTTAAGGATCTGAGGAAAAATGCTGGGCACTGGAAGCAGATTTGTTCTCAGG	1011

QY 1688 TCNTCAGACGACCTTTGGCNCNGCCCTGACGCGCANNNNNNNNNNNACGAACCTGANG 1747
DB 1014 TAGCGAAGAAACCTTGAGACAGAGGCTCAACGAGGAGACTTTTGCAGCGCTTTTGCG 1073
QY 1748 ACAAG 1753
DB 1074 AAAAG 1079

RESULT 13
US-09-252-991A-16475
Sequence 16475, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16475
LENGTH: 1548
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16475

Query Match
Best Local Similarity 5.3%; Score 131.6; DB 4; Length 1548;
Matches 572; Conservative 0; Mismatches 606; Indels 52; Gaps 5;

QY 471 CTTNGCCTCGAGCGGCTCGAGACTGCTCCGCTCGAGGTGCTCGCGGCGCAT 530
DB 336 CTTGTCCTCATGCGCGCGGCTGCGAGGCGCTGAGACGACCATGGTGGCGGTGTCAT 395
QY 531 GACCATCGCGCTCACTGCTGCTACGCGGCGGCGCATCTGCGGCTGCTGCTGCTGCA 590
DB 396 GGTGCTCAAGTGAACAGCTGGCGCGGCTGCTGCGGCTGCTGCGGCGCAAGGTGATGA 455
QY 591 GCGGCTCAACCACTTCTCAACCAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650
DB 456 CGCGCTGATCGGCTGATCAAGCGCGGCTGATCGGCTGATCGGCTGATCGGCTGATCG 515
QY 651 CTGCGCTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 710
DB 516 GCGTGTGCTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 575
QY 711 GGAANNCAAGTNCAGTNNNTCAAGAGGCGANNAGAAAGTCAATGNNCGCGCGGCGAGC 770
DB 576 CCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 617
QY 771 GATCGGCTCTTNGTCTCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 830
DB 618 GCTGGGCGGTGGCGGCTGGAAGCGCTGACCTGCGCGGCAAGGAGCGCTGCTGCTGCT 577
QY 831 CAACGCGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 890
DB 678 CAATGCGACCAAGTGTCTCAAGCGCTGAGCGGTGGCGGCTGCTGCTGCTGCTGCTGCT 737
QY 891 CTGCGCTCTGCGAGGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 950
DB 738 GTTCGCGCGCGCGCGCTGCTGCGGCGGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCG 797
QY 951 CTGCTTCAACCCNTTCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 1010
DB 798 GCGGCTTCAAGCGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCTCAAGCGCT 857
QY 1011 GCGCAACATCCGAGCGCTCTGAGAGGCGAGCANNTTTCCGCTCAACCAAGAGAGGT 1070

DB 858 GGCCTATCGCACTGCTCACCGCGAGGAGGTGCGGCTCCCATGAGGTGCGA 917
QY 1071 CAAGTCAAGACGACGAGGCTATTCGCCAGAGACCGCTACCCGCTCGGACGTGCGC 1130
DB 918 CAAGT-----CCAGACCCCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 953
QY 1131 TCAGTGTCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1190
DB 954 GCAGGTGATGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1013
QY 1191 GCGCGGAGTGCAGACCGGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1250
DB 1014 AGCC---AACGGGTGCTGCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1070
QY 1251 CCGCGGCACTTCCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1310
DB 1071 CCGCGGCACTTCCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1130
QY 1311 CCGCTGATCGGCAAGTCACTTCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1370
DB 1131 GCGCGAGATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1190
QY 1371 CCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1430
DB 1191 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1244
QY 1431 CCGTCACTTGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1490
DB 1245 CCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1304
QY 1491 GACCGAGTTCAGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1550
DB 1305 GCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
QY 1551 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1609
DB 1365 CCGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1424
QY 1610 ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1669
DB 1425 TGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484
QY 1670 ACCCGTNNNTCANNCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1699
DB 1485 AGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1514

RESULT 14
US-09-252-991A-16126/c
Sequence 16126, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 16126
LENGTH: 2295
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126

Query Match
Best Local Similarity 5.3%; Score 131.6; DB 4; Length 2295;
Matches 572; Conservative 0; Mismatches 606; Indels 52; Gaps 5;

QY 471 CTTGNGCCTGAGNGGCGCTCGAAGACTCGCTCCGCTCGAGGTCGTCGCGGCGCAT 530
 DB 1183 CTTGTCCTGAGCGCGCGCTCGGCGAGCGCTGAGCAAGCGCAAGGTCGCGCTGAT 1124
 QY 531 GACCATCGCGCTCACTCGCTCAAGCGCGGCGCTCGGCGGTCGCTGTCGTCGA 590
 DB 1123 GCTGCTCAAGGTGAAGAGCTGCGCGCGGCTGTCGCGCATCCCGCGCAAGGTGATGA 1064
 QY 591 GCGGCTCACCATTCTCTCAACGAGCATCAACCCCATGTCCTCCCTCGCGGACCAT 650
 DB 1063 CCGGCTGATCGCGCTGATCAAGCGCGAGGTCTATCGCAATCCCGCGAAGAGCTCGGT 1004
 QY 651 CTGCGGCTGAGGAGCGCTCTCCGCTGTCATCAACGCGCGCGCATCACCGGTCAACC 710
 DB 1003 GGGTGGCTCGGCGAGCTGGCGCGCTGAGCGCATGTCGCTGTCGTCGCGGAAAG 944
 QY 711 GGAACNNCAAGTNCACGTTNNTCACGAGGCGCANNGAAGATCATGTNCGCGCGAGGC 770
 DB 943 CCGGCGCGCGCATCGCGGTG-----AGTGGCTGCGCGCGCGCGCGAAGC 902
 QY 771 GATCGGCTCTTNGGCTCTGAGCGCGCTGCTCTCGCGCGCGAGAGGCTCTCGCTCTGT 830
 DB 901 GCTGGGCTGGCGCGGCTGAGCGCTGACCTTGGCGGAGAGAGGCTGCGCGCTGCT 842
 QY 831 CAAAGGACGCGCGCTCTCGCGCTCGATGCGCAACCTGCTCTGCAAGCGACACATGCT 890
 DB 841 CAATGGACACCAAGGTGTCCACCGCGCTACGCGGTGCGCGGCTGTTGAGGCGCGAAGCT 782
 QY 891 CTGGCTCTCTCGCGAGCGCTCAAGGCTCTTACGCTGAGGAGCATGTCGCGCGCG 950
 DB 781 GTTCGCGCGCGCGCGCTGCGCGGCTGCGCGCTGAGGCGCATGCTGCTGCGCGCG 722
 QY 951 CTGCTTCAACCTTCTCTCAAGCGCTCAAGCGCGCTCAACCGCGCATCGAGTCTGAGT 1010
 DB 721 GCGGCTTCAAGCGCGCTCAAGCGCGCTCAAGCGCGCTCAAGCGCGCTCAAGCGCG 662
 QY 1011 GCGCAACATCGAGCGCTCTCGAGGCGAGCGGCTGCTGCGCGCGCGCGCGAGT 1070
 DB 661 GGGCTTACGAGCATCTCTCAAGCGCGCGCGCGCGCGCGCTCTCTCAAGAGTGA 602
 QY 1071 CAAGGTCAAGAGAGAGAGGAGGATCTCGAGGAGCGCGTCAACCGCGCTCGCGAGTGGCC 1130
 DB 601 CAGAGT-----CAGAGCTCTTCTGCTGCTGCTGCGCGCGCGCGCGCGCGCG 566
 QY 1131 TCAAGTGGCTCGCGCGCTGTCAGCGCATGATTCAGCGCGCGCGCGCGCTGCTGCTG 1190
 DB 565 GCAAGTGGATGGCGCGCTGCTGACCGCATGCGCGCGCGCGCGCGCGCGCGCGCG 506
 QY 1191 GCGCGAGTGGAGCGAGCGAGCGCGCTGTCAGCGTTCAGAGCAAGAGAGCGCGCGCA 1250
 DB 505 AGCTC---AAGCGGCTTCCGACAACTCGGTGATTCGCGCGCGCGCGCGCGCGCG 449
 QY 1251 CCGCGGCACTTTCAGAGCGCGCGCTGTCAGAGCGAGTGAAGAGTCTGCGCTCGCTG 1310
 DB 448 CCGCGGCACTTTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 389
 QY 1311 CCGCTGATGCGAGAGTGAAGTTCAGCGCGCTCAAGCGAGTCTCAAGCGCGCGCGAG 1370
 DB 388 GCGCGAGATCGGTTCTGCTGTCGAGAGCGCGCATTCGCTGATGAGTGAAGTGAAGTGC 329
 QY 1371 CCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
 DB 328 GCAAGTTCGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 275
 QY 1431 CTTGCAATGTCGCGCGCGCGCGCATTCGAGGCTCGAGCACTTTCGCAACCGCGTAC 1490
 DB 274 CCGCGGAGTTCAGCG 215
 QY 1491 GACCGAGTTCAGCG 1550
 DB 214 GCTGACAGCTTCG 155

QY 1551 GCG 1609
 DB 154 CCGCAAAAGGCTTCG 95
 QY 1610 ACTGCGTCTCGAGCGCGCTGAGCTCCGCGCGATGAGTTCGAGTTCAGAGCAAGTTCG 1669
 DB 94 TGGCGCGCTGCGAGGCGCTGAGCTTCGCGCGAGGCGCTGAAGAGTTCGCGCAAGCTGAGC 35
 QY 1670 ACCGCTGCTGCG 1699
 DB 34 AGGCTGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5
 RESULT 15
 US-09-615-192A-97
 ; Sequence 97, Application US/09615192A
 ; Patent No. 6410718
 ; GENERAL INFORMATION:
 ; APPLICANT: Bloksberg, Leonard N.
 ; APPLICANT: Havukala, Ilka
 ; TITLE OF INVENTION: Materials and Methods for the
 ; TITLE OF INVENTION: Modification of Plant Lignin Content
 ; FILE REFERENCE: 11000.10034U
 ; CURRENT APPLICATION NUMBER: US/09/615,192A
 ; PRIOR FILING DATE: 2000-07-12
 ; PRIOR APPLICATION NUMBER: US 08/975,316
 ; PRIOR FILING DATE: 1997-11-21
 ; PRIOR APPLICATION NUMBER: US 08/713,000
 ; PRIOR FILING DATE: 1996-09-11
 ; PRIOR APPLICATION NUMBER: US 09/169,789
 ; PRIOR FILING DATE: 1998-10-09
 ; NUMBER OF SEQ ID NOS: 405
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO: 97
 ; LENGTH: 577
 ; TYPE: DNA
 ; ORGANISM: Pinus radiata
 US-09-615-192A-97
 Query Match 4.5%; Score 111; DB 4; Length 577;
 Best Local Similarity 54.0%; Pred. No. 2.8e-12;
 Matches 285; Conservative 0; Mismatches 234; Indels 9; Gaps 3;
 QY 786 TCTCAGCGCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 845
 DB 52 TCTGAAACCGTTTAAATGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 111
 QY 846 CTCGCGCTGATGAGGCGCGCTGCTGTCAGAGCGAGATGCTGCTGCTGCTGCA 905
 DB 112 GGGATCG 171
 QY 906 GCGGCTCAGCGCTCTTACGCTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 965
 DB 172 GATTCGCTGCGCGCTTCTGCGAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCG 231
 QY 966 CCGCAGCG 1025
 DB 232 AACCAACAGTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288
 QY 1026 GCTCTGAGGCG 1085
 DB 289 CTTCTCTCAAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345
 QY 1086 CAGGCGCATTCG 1145
 DB 346 GTTGAGCAAAACGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
 QY 1146 GCTGTCAGCGCATGATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1205
 DB 406 TCGATCAAGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 462
 QY 1206 GACCGCAACCG 1265

Db 463 CAACGACAATCCGTTAATGATGTCTCCAGGGACATGGCTCTCCACGGCGGCAATTCCA 522
QY 1266 GGGGNCCTGCTGTGGNACACGATGAGAAAGACTGCGCTCGGCTCGC 1313
Db 523 GGGAAACCCCATCGAGTTTCCATGACACATGCGAATCTCTTTGGC 570

Search completed: September 12, 2004, 04:26:57
Job time : 199.6 secs

Qy	2221	NNNCNNNNNNCTTTTGGNNNTCGNNTCNTGCNNNNCGGANTTNNCGNNNNNN	2280
	2221	NNNCNNNNNNNNCTTTTGGNNNTCGNNTCNTGCNNNNCGGANTTNNCGNNNNNN	2280
Db	2281	TNNNNCTNTTNCCTNNCTGCNNNNANCGNGTCNTTNNNNCTNNNGTNTNNNNNNC	2340
	2281	TNNNNCTNTTNCCTNNCTGCNNNNANCGNGTCNTTNNNNCTNNNGTNTNNNNNNC	2340
Qy	2341	NGTNNNCANNNACNCTNTNNNNNNANCGGANNGATNNAGNTTNCGNGNCNNNNNN	2400
	2341	NGTNNNCANNNACNCTNTNNNNNNANCGGANNGATNNAGNTTNCGNGNCNNNNNN	2400
Db	2401	NNNNANAAA 2408	
	2401	NNNNANAAA 2408	
Qy	2401	NNNNANAAA 2408	
	2401	NNNNANAAA 2408	

```

RESULT 2
US-09-939-408A-29
/ Sequence 29, Application US/09939408A
/ Patent No. US20020102712a1
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ APPLICANT: Kootstra, Anna
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
/ TITLE OF INVENTION: Using Same
/ FILE REFERENCE: 29479/500NSCA
/ CURRENT APPLICATION NUMBER: US/09/939,408A
/ CURRENT FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 09/624,693
/ PRIOR FILING DATE: 2000-07-24
/ PRIOR APPLICATION NUMBER: PCT/US01/23270
/ PRIOR FILING DATE: 2001-07-24
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 29
/ LENGTH: 2163
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Preferred
/ OTHER INFORMATION: theoretical sequence based in part on SEQ ID NO:20
/ NAME/KEY: CDS
/ LOCATION: (1)..(2163)
/ OTHER INFORMATION:
/ US-09-939-408A-29

```

Query Match	78.5%	Score 1944	DB 9	Length 2163
Best Local Similarity	93.2%	Pred. No. 0		
Matches 2037	Conservative	0	Mismatches 126	Indels 23
			Gaps	7
QY	1	ATGGCCCCCTTCNTCGACTGATCGGACCTCGTGTGGCCACGGGCTCNCMAACGNTNG	60	
Db	1	ATGGCCCCCTTCBTTGACTCGACTCGGACCTTCBTTGGCCACGGCCTDVAACGGVTG	60	
QY	61	CAGCGCCNNCCGNCNNNGCCNCAACGGAGCCCACTCCACNCTCNCNCGCCAGCGNG	120	
Db	61	CAGCGCCCTCGHCCAAATCGSCMAACGGAGCCCACTCTCACTCTCNCMGCCGCCBC-CGG	119	
QY	121	GCTCGCTCTCTCCCGACACCCAGNNGACGCAGCTCGACTTCGTGAGAGNAGTCTCTGCGCG	180	
Db	120	GCTCGCTCTCTCCCG-CCACCCAGVHAGAGCGCTCGACTGTGBAGVAAATCTTGCGCG	178	
QY	181	ACCCCAACCCNNAACGNAAGNNNTCGAACTCGACGGGTACACCTCACCTCGAGNACGTG	240	
Db	179	ACCCCAACCAAC---GACGMSVTGAACTCTGACGGGTACACCTCACCTCGHGAAGTGG	235	
QY	241	TGGCGCCCGCCGNCGGCAAGGGCGGCACNCTGTCGCGTTCNAGACAGTCACAGATCTCGCGCA	300	
Db	236	TGGCGCCCGCBGCGCAAGGGCGGCHBGTGCGCGTC-CAGAGAGTCGACGATCTCGCGCA	294	
QY	301	AAGATGCAACAAAGAGTGGAGTTCCTCCGANNCCAGCTCNCACCAAGAGTCTACGAGTTC	360	

Db	295	AAGATGCAAAVABGTTGAGTTTCTCCGBCACACTGBAACAACGBATCTACGGHCTC	354
OY	361	ACACATGGTTTGCGGCGCTCGGCGGACACACCGAGCTGAGGATGCNATCTCGCTCCAGAG	420
Db	355	ACGACCTGGTTTCGGGGGCTCGGCCGACACCGGACTGAGGATGCVATCTCGCTCCAGAG	414
OY	421	GCNCTCTTCGAGCACACGACTCTGCGGTGTCCTCCNACGTCGNTGANTCTTTCNGCTTC	480
Db	415	GCCTCTCTCGAACACACACTCTGGGGTGTCTCCBACGTCGATCGABCTCTTTCGCTTC	474
OY	481	GGCGCGGCGCTGGAACATCGCTTCGCTCGAGAGTGTCTCGCGCGCCATGACATCCGC	540
Db	475	GGHCGGGCGCTTGAAATCTCGCTTCGCTCGAAGTGTCTCGCGGCGCATGACATCCGC	534
OY	541	GTCACTCGCTCAACNCGCGGCACTCGGCGGTCCGCTCTGTGCTCTCGAGGCGCTCAC	600
Db	535	GTCACTCGCTCAACNCGCGGCACTCGGCGGTCCGCTCTGTGCTCTCGAGGCGCTCAC	594
OY	601	AACCTCCCTCAACAGGAGATACCCCCCATGTGCTCCCTCGGCGGACCATCTCGCGCTCG	660
Db	595	AACCTCCCTCAACAGGAGATACCCCCCATGTGCTCCCTCGGCGGACCATCTCGCGCTCG	654
OY	661	GGGACCTTCCCGCTCTCTCATGATCGCGCGGCATCAACGCGTACCCCGGACNNCAAG	720
Db	655	GGGACCTTCCCGCTCTCTCATGATCGCGCGGCATCAACGCGTACCCCGGACDBAAG	714
OY	721	GTNCAAGTNNTCAGAGAGGCAANNAGAAATCATGTNCGCCCGAGAGGCGATTCGCGCTC	780
Db	715	GTNCAAGTNNTCAGAGAGGCAANNAGAAATCATGTNCGCCCGAGAGGCGATTCGCGCTC	774
OY	781	TTNGCTCTGACCCCGTGTCTCGGCGCCGAAAGAGGATCTCGATCTCGTCAACGGCAG	840
Db	775	TTBGGCTCGAGCCCGTGTCTCGGCGCCGAAAGAGGATCTCGATCTCGTCAACGGCAG	834
OY	841	GCCGCTCTCGCTCGATGGGACCTTGCTCTGACAGAGCACATGCTCTCGCTCTC	900
Db	835	GCCGCTCTCGCTCGATGGGACCTTGCTCTGACAGAGCACATGCTCTCGCTCTC	894
OY	901	TGCGAGGGCTCAGGGCTCTNACGGTGGAGGCGCATGTGCGGACGCGCGCTCGTTCCAC	960
Db	895	TGCGAGGGCTCAGGGCTCTCTACGGTGGAGGCGCATGTGCGGACGCGCGCTCGTTCCAC	954
OY	961	CCNTTCTTCGACGAGCTCAGCGGCTCACCAGCCGAAATCGAGGTGCGCGCAATC	1020
Db	955	CCNTTCTTCGACGAGCTCAGCGGCTCACCAGCCGAAATCGAGGTGCGCGCAATC	1014
OY	1021	CGAGCGCTCTGAGAGGAGCGNNCTTGGCTGCAACAAGAGAGAGGTCAAGGTCAAG	1080
Db	1015	CGAGCGCTCTGAGAGGAGCGNNCTTGGCTGCAACAAGAGAGAGGTCAAGGTCAAG	1074
OY	1081	GACGAGAGGGCATTTCTCGGCAAGACCGGTACCGGCTCGGACGTCGCTCATGTGCTC	1140
Db	1075	GACGAGAGGGCATTTCTCGGCAAGACCGGTACCGGCTCGGACGTCGCTCATGTGCTC	1134
OY	1141	GGCGCGCTGTGAGGAGATGATTCAGGCCAAGCNGTCTCTCGTCAAGGCGGNCAG	1200
Db	1135	GGCGCGCTGTGAGGAGATGATTCAGGCCAAGCNGTCTCTCGTCAAGGCGG--AG	1194
OY	1201	TGACACACCGAACACCGGCTCATGAGCGTGAAGAACAGNACCCACACACGCGCGCAAC	1260
Db	1192	TGACACACCGAACACCGGCTCATGAGCGTGAAGAACAGNACCCACACACGCGCGCAAC	1254
OY	1261	TTTCAGAGCGNCGGCTGTGCAAAACAGATGGAAGATCTGCTCGGCTCGGCTTGAAT	1320
Db	1252	TTTCAGAGCGNCGGCTGTGCAAAACAGATGGAAGATCTGCTCGGCTCGGCTTGAAT	1314
OY	1321	GGCAACCTCAATTACGCGAGCTCACCGAATGCTCAAGCGCGGATGAACCGCGGCTTN	1380
Db	1312	GGCAACCTCAATTACGCGAGCTCACCGAATGCTCAAGCGCGGATGAACCGCGGCTTB	1374
OY	1381	CCNTCTGCTCGCTGCGGAGGACCCNTGCTCTCTTATCATGCAAGGCGCTCGACAT	1440

Db 1372 CCBTCTGCTGCTGCGGAGACCCBFGCTCTCTATCACTGAGAGGCTCGACATT 1431
QY 1441 GCCGCGNGCGNCTACCTTGGAGCTGGNCACTTTGCCAACCCGGTACGACCGACGTC 1500
Db 1432 GCCGCBGCBGCBTACCTTGGAGCTGGHCACTTTGCCAACCCGGTACGACCGACGTC 1491
QY 1501 CAGCGCGNAGATGGGCAACGAGCGCTCAACTGCTGCTGATCATCTCGGCGCGCCG 1560
Db 1492 CAGCGCGNAGATGGGCAACGAGCGCTCAACTGCTGCTGATCATCTCGGCGCGCCG 1551
QY 1561 ACNCGCGAGGCAACGAGCTCTTCTCTCTCTCTGCGCCACCTTACTGCTGCTGCTC 1620
Db 1552 ACBGCAGAGGCAACGAGCTCTTCTCTCTCTCTGCGCCACCTTACTGCTGCTGCTC 1611
QY 1621 CAGCGCGTCACTCGCGCGATGGAGTTCGAGTTCAGAGACCACTTGCACCGTNTTC 1680
Db 1612 CAGCGCGTCACTCGCGCGATGGAGTTCGAGTTCAGAGACCACTTGCACCGTNTTC 1671
QY 1681 NCNCGCTCTNAGAGAGACTTTGGCNCGCCCTCGACGCGNNNNNNNNNNNNNACAA 1740
Db 1672 VCBDCGCTGCTVAGAGAGACTTTGGCTCCTGCTCGACGCG-----MACGAA 1719
QY 1741 CTGNGGACAAAGTCAACAAAGCTGACAAAGCTGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 1720 CTGNGGACAAAGTCAACAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1779
QY 1801 CTGAGGCGGCTGAG 1860
Db 1780 CTGAGGCGGCTGAG 1839
QY 1861 NNCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 1840 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1897
QY 1921 CCGCGGAG 1980
Db 1898 CCGCGGAG 1957
QY 1981 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
Db 1958 CGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2017
QY 2041 TCCGCGAG 2100
Db 2018 TCCGCGAG 2077
QY 2101 TGAAGATCGGACCAAGCTCTCGCGATCTAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2078 TGAAGATCGGACCAAGCTCTCGCGATCTAGAGAGAGAGAGAGAGAGAGAGAG 2137
QY 2161 ACGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2186
Db 2138 ACGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2163

RESULT 3
US-09-939-408a-18
Sequence 18, Application US/09939408A
Patent No. US20020102712A1
GENERAL INFORMATION:
APPLICANT: Yoshida, Roberta
APPLICANT: Kocistara, Anna
TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
FILE REFERENCE: 29479/500NSCA
CURRENT APPLICATION NUMBER: US/09/939,408A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 09/624,693
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: PCT/US01/23270
PRIOR FILING DATE: 2001-07-24
NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 2439
TYPE: DNA
ORGANISM: Rhodotorula toruloides
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(2148)
US-09-939-408a-18

Query Match 68.7%; Score 1699.4; -DB 9; Length 2439;
Best Local Similarity 79.4%; Pred. No. 0;
Matches 1913; Conservative 0; Mismatches 459; Indels 36; Gaps 6;

QY 1 AAGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
Db 1 AAGGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
QY 61 CAGCGCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCGNCG 120
Db 61 CAGCGCTG-----CATGCGCTGCGACCAACTC-----GAGTGCAG 100
QY 121 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db 101 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 160
QY 181 ACCCGACCGNAG 240
Db 161 GCGCGACCGAGCTGCA-----GCTGCAACTGCAAGGCTGCTGCTGCAACTGCGAGAGCTG 217
QY 241 TGGGCGCGCGNCGAG 300
Db 218 TCTGCGCGCGAG 276
QY 301 AAGATCGACAAAG 360
Db 277 AAGATCGACAAAG 336
QY 361 ACGAGTGGTTTCGCGGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 337 ACGAGTGGATTTGGGAG 396
QY 421 GCTGCTGCTGAG 480
Db 397 GCTGCTGCTGAG 456
QY 481 GAGCGGAGCTGAG 540
Db 457 GAGCGGAGCTGAG 516
QY 541 GTCAACTGCTGACNCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Db 517 GTCAACTGCTGACNCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 576
QY 601 AACTTCTCAACAG 660
Db 577 AACTTCTCAACAG 636
QY 661 GAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 637 GAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 696
QY 721 GTNCGAGNNNNNCGAG 780
Db 697 GTNCGAGNNNNNCGAG 756
QY 781 TTNGTCTGAG 840
Db 757 TTCAACTCTGAG 816
QY 841 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Db 817 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 876

[illegible][illegible]

D	3	ATGAGCCCTTCTTGGATCGCTGCGCACACGCTGACCAAGGCTTTTACCAAGGCTCG	96
Q	61	CAGCCGANNCCGNCNNNGNCNCAACGGGGCCACGTCACNTCTGNGNGGCCNAGCG	120
D	97	CAGCGCGCTCCAGCAAGTCGAGCTCGGGGGCCCACTTCGCTTCGGCGCAGCG--CTG	154
Q	121	GCTCGCTCTCCCGACACCCAGNNAGCGAGCTCGACTCGTACGTNAGNAGATCTCGCG	180
D	155	GCTCGATGCGCACCGCGCGACAGCTGCAAGCTCGAGATCGTGCAGAGAGCTCTCAAG	214
Q	181	ACCCCAACGNNACGNAAGNNTTGAACTCGACGGGTACCTTCACCTTCGGNACGTCG	240
D	215	ACCCCAACCCAC--GACTCGTGAGGTCACAGCGGTACACCTCCACCGTCGATGAGTTG	271
Q	241	TGGGGCCGCMCGAAGGGCCGNCNNGTCGCGCGNCAGACAGNCAGACAGATCCGCGCA	300
D	272	TGGGCGCCCGCGCAAGGGGCGCAGGGTTCGCGTC-CAGAACGACACAGATCCGCGCA	330
Q	301	AAGATGACANAGNGTCGAGTTCCTCGNNCNCAGCTCAACAAAGNTCTAAGSNGTC	360
D	331	CGCGTCGACAAAGCGTCGACTTCCTCAAGGCCAGCTTCAGAACTCGCTTACGSAATC	390
Q	361	ACGACTGGTTTGGGGGCTCGGCGGACACCCGATCTAAGATTCMACTTCGTTCCAGAG	420
D	391	ACCAAGGTTTGGTGGCTCGAGCGACACGAGACTAAGATCACTAGCTCCAGAG	450
Q	421	GCNCTCTTGACACACCACTCTGGCGGTNTCTCCNACGTCGATTCAMTCTTTCGCGTC	480
D	451	GCGCTCATGACACACGATCTGGGCGGTGACCGCAGCGTCGCGTCTTCCTTACGGCTC	510
Q	481	GCNCGGCGCTCGAAGACTCGCTTCGTCGAGGTCTCTCGGGCGCCATGACATCCGC	540
D	511	GGACGGGCGCTGAAACAACGCTTCGCTGAGGTCTCTCGCGCCCATGCTACCTCGC	570
Q	541	GTCAACTGCTCAGNCGGGGCACTCGGNGTCGCGCTCGTCTCTCGAAGCGCTCACG	600
D	571	GTCAACTGCTCAGCGGTGCGACTCGGCGGTCCGCTCGTCTTTAAGGCGCTCACG	630
Q	601	AACCTTCTCAACACAGGAGTACACCCCATGTCCTCCCTCGCGGACCATCTCGGCGTCG	660
D	631	AACCTTGAACACACGATCACGCCCATGTCCTCCCTCGGCGCTCATCTCGGCGTCG	690
Q	661	GCGGACCTCTCCCNCTCTCTAATATGCGCGCGCCATACCGGTACACCCGGAACNNCAG	720
D	691	GCGGACCTCAGCGCGCTCTGTATATGCGCGCGCATACCGGTACACCCGACGTCAG	750
Q	721	GTNACAGTNNTCAGAGGCGANNAGAGATCATGTNCGCCCGCAGCGCATCTCGCTC	780
D	751	GTTCAAGTTTGCAGAGGAACCGAAGATCATGTTGGCGGACGSCATCTCGCTC	810
Q	781	TTNAGTCTCGAGCCGCTGTCTTGCGCCGGAAGAGGTTTCGCTCTGTCAACGCGACG	840
D	811	TTTGGTCTGAGGAGTGTCTCTCGCGCCGGAAGAGGTTTCGCTCTGTCAACGGAACG	870
Q	841	GCGGTTCGCGTATGAGCAACCTCGCTCTGACAGACACATNCTCTCGCTCTC	900
D	871	GCGGTTCGCGCTCATGAGCACTCTGATGTGACGACTCGACATGCTCTCGCTCTC	930
Q	901	TGCGAGGCGCTCAGGACTCTNACGAGTCGAGCGCATGGTCGCGCACCGCGCTGTTCCAC	960
D	931	TGCGAGGCGCTTACGAGCTCTCAGAGTGAAGCCATGGTGGCGCAGAGGCTCTGTTCCG	990
Q	961	CCNTTCCTCAAGAGTCAAGCGGCTTCACCCGACCCAGATGGAAGTTCGGGGCAACATC	1020
D	991	CGGTTATCAAGAGTCTCGCGCCGCAACCCGCGAGGTGAGGTCGGGCGCAATC	1050
Q	1021	CGCAGCTCTCGAGGCGAGCNGTTTTCGCTCCACACGAGGAGAGGTCAAGTCAAG	1080
D	1051	CGCAGCTCTTTCGGGCTCGTCTGTTTCGTTGAGCAGAGGAGGCTCAAGGCTCAAG	1110
Q	1081	GAGGAGAGGAGATCTCTCGGCAAGACCGCTACCCGCTTCGCACTCGGCTCAGTGGGCTC	1140
D	1111	GAGGAGAGGAGATCTTTCGCGAGACCGCTACCCGCTTCGCACTCGGCTCAGTGGCTC	1170

QY 2221 NNNCCNNNNNNCTTTNNNNNTGNNNTGNNNNNNNNNNCGGANNNTNNNNNNNNNN 2280
Db 2233 ACCCCAAGACCAAGCTTTTCCAGCGCTGTGTGTCAGCAAGACGACTTCTTCATACACA 2292
QY 2281 TNNNNCTNNCTNNCTCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
Db 2293 TGTGCTTACTCTCTGCGCGGTGATCATGCTCTGAGTCTTCTGATCCGCGTCTCTC 2352
QY 2341 NGTNNNNANNACNNCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2400
Db 2353 GTGTGCTCA-GTACACGCTATATAGAGCGCTGAAATGATGCAAGTCTTCAAAAAA 2411
QY 2401 NNNNNAAA 2408
Db 2412 AAAAAAAA 2419

RESULT 5
US-09-765-873A-9
; Sequence 9, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; CURRENT FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
US-09-765-873A-9

Query Match 65.7%; Score 1626.4; DB 9; Length 2151;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 1829; Conservative 0; Mismatches 322; Indels 35; Gaps 5;

QY 1 ATGAGCCCTCTCCTGACTGATCGGACCTCGANTGCGCAAGGCGTCCNNACGGANTG 60
Db 1 ATGAGCACTCTCGCTCGACTGATCTGCACTGCTTGGCAAAGCGGCTGGCATCCGAAAG 60
QY 61 CAGCGCGNNCCGNCNNNNCGNCNACGAGCGCCGACGCTCCACNCTCNGCGCGCGNG 120
Db 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GAGTCGCGAG 100
QY 121 GCTGCTCTCCCGGACCGACCGAGNNAGCGAGCTGAGCATGCTNNAGNAGATCTTCGCG 180
Db 101 GCTGCACTCTGCGCAACCCAGGTGACGAGGTGACATGCTCAAGAAATGCTCGCG 160
QY 181 ACCCAACCGNACGAGNNNTGAACTGACGAGGTACACCTCTCACTCCGAGNGAGCTCG 240
Db 161 CGCGACCGACTGCA---CGCTGAACTGACGAGGTACTCGCTCAACTCGGAGAGCTCG 217
QY 241 TCGGCGCGCGCGCAAGGGCGCGNCNGTCTCGCTCNCAGACAGNAGAGAGATCCGCGCA 300
Db 218 TCTGCGCGCGGAGAAAGGGGAGGCGCTGTCTCGGCTCAAGAGACAG-GCAGAGATCCGCTCA 276
QY 301 AAGATGACANAGNGTGAAGTCTCCGNNCCAGCTCNAACAGAGTCAACGAGATC 360
Db 277 AAGATTGACAAATCGGTGAGTCTTTCGCGCTGCAACTTCATAGAGGTCTACGCGCTC 336
QY 361 AGCAGTGTGTTGCGCGGCTCGCGCGACACCGGAGTGAAGATGCAATCTCGCTCGAAG 420
Db 337 AGCAGTGTGATTTGCGCGAGATCCGAGACACCGGACCGAGAGCGCATCTCGCTCGAAG 396

QY 421 GCTCTCTTGAGACCAAGCTTTCGCGGTGTTCTCCNAGCTGANTGANTCTTCCGCTC 480
Db 397 GCTCTCTTGAGACCAAGCTTTCGCGGTGTTCTCCNAGCTGANTGANTCTTCCGCTC 456
QY 481 GAGCGGCGCTCGAAGACTCGCTTCGCTCGAGGTGTCGCGGCGGCGCATGACATCCG 540
Db 457 GCGCGGCTCTCGAAGACTCGCTTCGCTCGAGGTGTCGCGGCGGCGCATGACATCCG 516
QY 541 GTCACTGCTCAACNCGCGGCACTCGCGNGTCCGCTCTGTGCTCTTGAGGCTCAC 600
Db 517 GTCAAGCTTGACCGCGGCACTCGCGTCCGCTCTGTGCTCTTGAGGCTCAC 576
QY 601 AACTTCTCAACCGAGCATCAACCCCATCTGTCCTCCGCGGCGCATGCTCGCGCTG 660
Db 577 AACTTCTCAACCGAGCATCAACCCCATCTGTCCTCCGCGGCGCATGCTCGCGCTG 636
QY 661 GCGACCTCTCCCNCTCTNTACATCGCGCGGCACTACCGGTCAACCGGACNNCAAG 720
Db 637 GCGACCTCTCTCTCTCTCTCTACATTTGAGCGGCACTACCGGTCAACCGGACGCAAG 696
QY 721 GTNCACTNNNNNCAAGAGGCGGANNAGAAAGATCATGTCGCGCGGAGCGATCGCTC 780
Db 697 GTNCACTGTGTCAAGAGGCGGAAAGAAAGATCTGTACGCGCGGAGCGATGCGCTC 756
QY 781 TTNGGTCTGAGCGCGCTGCTCTCGCGCGGAAAGAGGCTCTGCTGCTCAACGCGCAG 840
Db 757 TTCAACTGAGCGCGCTGCTCTCGCGCGGAAAGAGGCTCTGCTGCTCAACGCGCAG 816
QY 841 GCGGTCTCGGCTGATGAGCGACCTCTGCTGCAAGCGACACATGCTCTGCTCTC 900
Db 817 GCGGTCTGAGCATGATGAGCGACCTCTGCTGCAAGCGACACATGCTCTGCTCTC 876
QY 901 TCGAGAGGCTCAAGCGCTCTNAGCGNNAGGCGCATGTCGCGCAAGCGCGCTCTCCAC 960
Db 877 TCGAGTCTCTCAAGCGCATGACGCTGAGCGATGTCGCGCGCGCTCTCTCTCAC 936
QY 961 CCNTTCTCAAGCGCTCAAGCGCGCTCAACCGGACCGAGTGAAGTGGCGCGCAATC 1020
Db 937 CCTTCTCTCAAGCGCTCAAGCGCGCTCAACCGGACCGAGTGAAGTGGCGCGCAATC 996
QY 1021 CGAGCTCTCTGAGAGGCGGNNGTTTCCGCTGCAACCGAGAGAGAGTGAAGTCAAG 1080
Db 997 CGAAGCTCTGAGAGGCGGCTTGTGTGCAACATGAGAGAGTGAAGTCAAG 1056
QY 1081 GAGAGAGAGGATCTCTCGCGAGAGCGCTACCGGCTCCGCAAGTCAAGTCAAGTCTC 1140
Db 1057 GAGAGAGAGGATCTCTCGCGAGAGCGCTACCGGCTCCGCAAGTCTCTCAAGTCTC 1116
QY 1141 GGTGCTCTGTAAGCATGATTAACGCGCAAGCGNGTCTCTGCTGAGGCGCGNAG 1200
Db 1117 GGTGCTCTGTAAGCATGATTAACGCGCAAGCGCGCTCTCAAGTCAAGTCAAGTCAAG 1176
QY 1201 TCGAGAGCGCAACCGGCTCTGTAAGCGTCAAGCAAGAGAACCAAGAGCGGCGCAAC 1260
Db 1177 TCGAGAGCGCAACCGGCTCTGTAAGCGTCAAGCAAGAGAACCAAGAGCGGCGCAAT 1236
QY 1261 TTCAAGCGNNCGCTGTGCAACAGATGAGAGATCTGCTGCTGCTGCTGCTGCTGATC 1320
Db 1237 TTCAAGCGNNCGCTGTGCAACAGATGAGAGATCTGCTGCTGCTGCTGCTGCTGATC 1296
QY 1321 GGTAACTCACTTCAAGAGTCTCAAGAGATGCTCAAGCGCGGATGAACCGGCGCTN 1380
Db 1297 GGTAACTCACTTCAAGAGTCTCAAGAGATGCTCAAGCGCGGATGAACCGGCGCTC 1356
QY 1381 CCNTCTGCTGCTGCTGCGAGGACCTNTGCTCTCTCTACTGCAAGGCTCTGCACTT 1440
Db 1357 CCTCTGCTGCTGCGGCGGAGAACCTCTGCTCTCTCTCTCTCTCTGAGGCTCTGCACTC 1416
QY 1441 GCGCGNCGNNTAATCTTGAAGTCTGAGNACCTTTCGCAACCGGCTNACGACCAAGTCTC 1500
Db 1417 GCGCGTGGCGGCTAGACCTCGAGTGTGGAGACCTCTGCGCAACCTGTGTGACGACGATGTC 1476

[illegible]

Db 101 GTGTGACCTGCTCCCAACCCAGGTCAGGAGGTGAGATGTGGAAMAATGTGTCGG 160
 Qy 181 ACCCCACCGNNAACGACGNNNTTGAACCTGACGGGTAACCCCTCACCCCTCGAGAGCTCG 240
 Db 161 GCGCCACGACCTCGA---CGCTCGAAGCTGAGGGTACTGTGCTCAACTCGAGAGAGCTCG 217
 Qy 241 TCGGAGCGCGGCGCAAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 Db 218 TCTCGGCGCGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 276
 Qy 301 AAGATCGACCAAAAGAGTGAAGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
 Db 277 AAGATTGACAAATCGGTTGAGTTCTTGCGCTCGAACTCTCCATGAGCGTCTACGGGCTC 336
 Qy 361 ACGATGCTGTTGCGGCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
 Db 337 ACGATGAGTTTGGGCGGATCGGCGAGACCCGCGACCGGAGGAGCGCATCTCGCTCGAAG 396
 Qy 421 GCGTCTCTGAGACCAAGCTCTGCGGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
 Db 397 GCTCTCTCGACACAGCTCTGCGGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456
 Qy 481 GAGCGGCGGCTCGAGAACTCGCTCGGCTCGAGGCTGTCGCGGCGGCGGCGGCGGCGGCGG 540
 Db 457 GCGCGGCGGCTCGAGAACTCGCTCGGCTCGAGGCTGTCGCGGCGGCGGCGGCGGCGGCGG 516
 Qy 541 GTCAACTCGCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
 Db 517 GTCAACAGCTTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 576
 Qy 601 AACTTCTCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 Db 577 AACTTCTCTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
 Qy 661 GCGGACCTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 Db 637 GCGGACCT 696
 Qy 721 GTTACGCTGTTTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
 Db 697 GTTACGCTGTTTACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756
 Qy 781 TTGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
 Db 757 TTGAGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 816
 Qy 841 GCGGCTCTCGGCTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
 Db 817 GCGGCTCTCGGCTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 876
 Qy 901 TCGGAGGCGGCTCGGCTCTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
 Db 877 TCGGAGTCTGCTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 936
 Qy 961 CCGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
 Db 937 CCGTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 996
 Qy 1021 CGGAGCTCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
 Db 997 CGGAGCTCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1056
 Qy 1081 GAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
 Db 1057 GAGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1116
 Qy 1141 GCGCGGCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
 Db 1117 GCGCGGCTCTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
 Qy 1201 TCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260

Db 1177 TCGAGAGCCGACCAACCTCTCTATCGAGCTGAGAGAAAGACTTTCGACGAGGCGGCAAT 1236
 Qy 1261 TTCCAGCGGAGCGGCTGTGCGGCAACAGATGAGAGAAAGCTGCGCTCGGCGGCGGCGGCGG 1320
 Db 1237 TTCCAGGCTGTGCGGCTGTGCGGCAACAGATGAGAGAAAGCTGCGCTCGGCGGCGGCGG 1296
 Qy 1321 GCGAAGCTCACTTCAAGGAGCTCACGAGATGCTCAAGCGGCGGCGGCGGCGGCGGCGGCGG 1380
 Db 1297 GCGAAGCTCACTTCAAGGAGCTCACGAGATGCTCAAGCGGCGGCGGCGGCGGCGGCGG 1356
 Qy 1381 CCGTCTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
 Db 1357 CCGTCTCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1416
 Qy 1441 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
 Db 1417 GCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1476
 Qy 1501 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
 Db 1477 CAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1536
 Qy 1561 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1620
 Db 1537 ACGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1596
 Qy 1621 CAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1680
 Db 1597 CAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1656
 Qy 1681 NCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1740
 Db 1657 GCTGCTCTATGAGCAAGCACTTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1716
 Qy 1741 CTGAGGCAAGAGTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
 Db 1717 CTGAGGCAAGAGTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1776
 Qy 1801 CTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
 Db 1777 CTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1836
 Qy 1861 NNGTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
 Db 1837 TCGTACGCTC-----GCTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1885
 Qy 1921 CCGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
 Db 1886 CCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1945
 Qy 1981 GGTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
 Db 1946 GGTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2005
 Qy 2041 TCGGCGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
 Db 2006 TCGGCGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2065
 Qy 2101 TGAAGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
 Db 2066 TGAAGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2125
 Qy 2161 ACGTCTCTGCAAGATGCTGCGGCTAG 2186
 Db 2126 ACGTCTCTCTCAAGATGCTGCGGCTAG 2151

RESULT 9
 US-10-374-366-9
 ; Sequence 9, Application US/10374366
 ; Publication No. US20040014085A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Xiao-song

APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: CA1794 US NA
CURRENT APPLICATION NUMBER: US/10/374,366
PRIOR FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/350,279
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 9
LENGTH: 2151
TYPE: DNA
ORGANISM: Rhodospiridium glutinis
US-10-374-366-9

Query Match 65.6%; Score 1624.8; DB 16; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1826; Conservative 0; Mismatches 323; Indels 35; Gaps 5;

QY 1 ATGACCCCTCCCTCGATCGATCGGACCTCGNTGCCAACGGGNTCCNAACGGTNG 60
DB 1 ATGACCCCTCCCTCGATCGATCGGACCTCGNTGCCAACGGGNTCCNAACGGTNG 60
QY 61 CAGCGCTGT-----CAATGGGCTCGACCACTC-----GCACTCGAG 100
DB 61 CAGCGCTGT-----CAATGGGCTCGACCACTC-----GCACTCGAG 100
QY 121 GCTGCTCTCCGACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
DB 101 GCTGCTCTCCGACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 160
QY 181 ACCCCACCCGNNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAG 240
DB 161 CGCGACCGACTGGA---CGCTGAACTGAGCGGCTGCTCAACCTCGAGAGCTG 217
QY 241 TCGGCGCGCGCCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 218 TCTCGCGCGCGAG 276
QY 301 AAGATGCAAAAGAGTGAATCTCTCGGNNCAGCTTCAACAGAGTCTAGCGGTC 360
DB 277 AAGATGCAAAATCGATCGAGTCTTTCGCTCGCAACTCTCAAGAGCTGACGGCTC 336
QY 361 AACACTGGTTGGCGCGCTCGCGCGAGCAGCGGAGTGAAGATGATTCCTCCAGAG 420
DB 337 AACACTGGATTTGGCGGATCGAGACACCGGAGCAGAGAGCCATCTCTCCAGAG 396
QY 421 GCNCTCTCGAGCAGCAGCTGCGGTGTCCTCCNAGCTGAGTCCCTTNGCTC 480
DB 397 GCTCTCTCGAGCAGCAGCTGCGGTGTCCTCTGCTGCTGAGTCTGCTCCGCTC 456
QY 481 GAGCGGCGCTCGAGAACTCGCTTCGAGTTCGCGGAGCGGAGCGATCAATCCG 540
DB 457 GAGCGGCGCTCGAGAACTCGCTTCGAGTTCGCGGAGCGGAGCGATCAATCCG 516
QY 541 GTCAACTCGCTCAGCGCGGCACTCGGCGTCCGCTCGTCTCGAGCGCTGACC 600
DB 517 GTCAAGAGCTTGAACCGCGGCACTCGGCTGTCGCTGTCGTCGAGGCGCTTACC 576
QY 601 AACTTCTCAACCAAGGATCAACCCCATGTCCTCCCTCGGAGCAGCATTCGCGTCG 660
DB 577 AACTTCTCAACCAAGGATCAACCCCATGTCCTCCCTCGGAGCAGCATTCGCGTCG 636
QY 661 GAGGAGCTCTCCGCTCTGNTACATCGCGCGGAGTCAACCGGTCAACCGGAGCAG 720
DB 637 GAGGAGCT 696
QY 721 GTNACGTTNNACGAGGAGCANNAGAGATCATGTNCGCCGAGGAGCATGCGCTC 780
DB 697 GTGCACTGTCGACGAGGAGCAAGAGAAAGATCTGTACGCGCGGAGGAGCATGCGCTC 756
QY 781 TTNGGCTCGAGCGCGCTGCTCTCGCGCGGAGAGAGGCTCTGCTCTGTCACAGGACG 840

DB 757 TTCAACTCGAGCGCGCTGCTCTCGCGCGGAGAGAGGCTCTGCTCTGTCACAGGAC 816
QY 841 GCGCTTCGCGCTTCATAGGAGAGACCTGCTCTGACAGAGAGACAGAGTCTGCTCTC 900
DB 817 GCGCTTCGAGATGAG 876
QY 901 TCGAGCGCTCAAGGCTCTNAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 877 TCGAGCTGCTCAAGGAG 936
QY 961 CCNTTCTCAAGAGAGTCAAGCGGCTCAACCGGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 937 CCNTTCTCAAGAGAGTCAAGCGGCTCAACCGGAGAGAGAGAGAGAGAGAGAGAGAG 996
QY 1021 CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
DB 997 CGCAGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1056
QY 1081 GACGAGAGGAGGATTCGCGCAGAGCGGTAACCGCTTCGAGAGTCTGAGTCTC 1140
DB 1057 GACGAGAGGAGGATTCGCGCAGAGCGGTAACCGCTTCGAGAGTCTGAGTCTC 1116
QY 1141 GAGCGGCTGAGAGGAGATGATTCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB 1117 GAGCGGCTGAGAGGAGCTTATTCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1176
QY 1201 TCGAGACCGAGCAACCGCTCATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1177 TCGAGACCGAGCAACCGCTCATGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1236
QY 1261 TTTCAGGAGGAGGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1237 TTTCAGGAGGAGGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
QY 1321 GAGGAGCTCAACTTACGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1297 GAGGAGCTCAACTTACGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
QY 1381 CCNTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1357 CCNTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1416
QY 1441 GCGGAGGAGGAGTGAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1417 GCGGAGGAGGAGTGAAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
QY 1501 CAGCGGAGGAGATGAG 1560
DB 1477 CAGCGGAGGAGATGAG 1536
QY 1561 AGGCGGAGGAG 1620
DB 1537 AGGCGGAGGAG 1596
QY 1621 CAGCGGCTGAGCTCGCGGAGATGAGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAG 1680
DB 1597 CAGCGGCTGAGCTCGCGGAGATGAGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAG 1656
QY 1681 NCGGCTCAGTCAAG 1740
DB 1657 GCTGCTGATGAG 1716
QY 1741 CTGAG 1800
DB 1717 CTGAG 1776
QY 1801 CTGAG 1860
DB 1777 CTGAG 1836
QY 1861 NNGTCTGCGGAGGAG 1920
DB 1837 TCGTGAAGT-----GCTCTGCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1885

QY 1921 CCGCGAGAGGCGATCTTGGCTGACGCGGANNGTCCGGACNCCTTCTGGNCGGNCCTT 1980
| | | | |
Db 1886 CCGCCCAAGTCGGCGCATCTGCTGACCCCGCAAGTCCGGAGACCTTCTGGGCCCGCGT 1945
| | | | |
QY 1981 CGCGCTGCTGCGCGCGGCTCNCGTACTCTGTGCGCGGACCGGCGGCTCTGACTGTTTCG 2040
| | | | |
Db 1946 CGACCTCTGCGCGCGGCTCTGTACTCTGTGCGCGGACCTCAATCTCTTACGCCCTTCG 2005
| | | | |
QY 2041 TCCGCGAGAGTCCGCGCTCAAGGCCCCCGCGCGGACGCTTCTCTCGGCAAGAGAGG 2100
| | | | |
Db 2006 TCCGCGAGAGTCTGGCGCTCAAGGCCCCCGCGGAGAGTCTTCTCTCGGCAAGAGAGG 2065
| | | | |
QY 2101 TCACGATTCGCGACCAAGCTTCCCGATCTTACAGGCGCATCAAGNCGGCGCATCAAC 2160
| | | | |
Db 2066 TCGCGATTCGCGTCAAGCTTCTCAAGATCTACAGGCGCATCAAGTCCGGCGAGATCAACA 2125
| | | | |
QY 2161 ACCTCTCTCTCAAGATGCTCGCTAG 2186
| | | | |
Db 2126 ACCTCTCTCTCAAGATGCTCGCTAG 2151
| | | | |

RESULT 10
US-10-374-366-13
; Sequence 13, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: CL1794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; PRIOR FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 13
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-13

Query Match 65.6%; Score 1624.8; DB 16; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1828; Conservative 0; Mismatches 323; Indels 35; Gaps 5;

QY 1 AMGGCCCCCTCCNTGACTCGATCGGACCTCGANTCGCAAGGCGTNCMAAGGNTNG 60
| | | | |
Db 1 ATGGCACCCCTCGCTGACTGATCTGCACTGTTGCAAAAGGCGTGCATCCGCAAG 60
| | | | |
QY 61 CAGCGCCGNNCCGNCNNNGNCGNACGAGCGCCAGATCCACNCTCNGNCGGCCGNCNG 120
| | | | |
Db 61 CAGGCTGT-----CAATGGCGCTCGACCAACCTC-----GAGTGGAG 100
| | | | |
QY 121 GCTGCTCTCTCCGACACCAACCCAGNNAGCCAGCTGACATGCTGNAGAAATCTCTGCG 180
| | | | |
Db 101 GCTGCACTCTGCCCAACCAAGGTCAAGAGTGCATGCTGCAAGATGCTTGGCG 160
| | | | |
QY 181 ACCCGACCGNNACGNAAGNNNTGAACTCGAGGGGTACACCTTCACTCGGAGAGCTCG 240
| | | | |
Db 161 CGCGGACCGACTGA---CGCTGAACTCGAGGGCTACTCGTCAACTCGGAGAGCGTGG 217
| | | | |
QY 241 TCGGCGCGCGCGCAAGGCGCGCNCNGTCCGCGTNCAGACAGNCGAGAGATCCGCGCA 300
| | | | |
Db 218 TCTCGCGCGCGAAGAGGCGAGGCGCTGTCGCGGTCAAGAGACAG- GAGAGAGATCCGCTCA 276
| | | | |
QY 301 AAGATGACAAANAGNTCGAGTTCCTCCGNNCAGCTGNAACAAGAGTCTACAGGANTC 360
| | | | |
Db 277 AAGATTGACAAATCGGTGAGGTTCTTGCGCTGCAACTCTCATAGAGGCTTACGGCGTC 336
| | | | |
QY 361 ACGACTGCTTCCGCGCGCTCGCGCGACACCGAGCTGAGAGATGCAATCGCTCCAGAG 420
| | | | |

Db 337 AGGACTGATTTGGGGATTCGGCAGACACCCGCAACCGAGAGCGGCATCTCGCTCCAGAG 336
| | | | |
QY 421 GGNCTCTTGAGACACAGCTCTGCGGTGTNCTCCNAAGTCAANTCGTTCNGCCTC 480
| | | | |
Db 397 GCTCTCTTGAGACACAGCTCTGCGGTGTNCTCCNAAGTCAANTCGTTCNGCCTC 456
| | | | |
QY 481 GGNCGCGGCTTGAGAACTGCTTCCGCTGAGAGTCCGCGGAGCGCATGACATCCGC 540
| | | | |
Db 457 GGCAGGCTCTGAGAACTGCTTCCGCTGAGAGTCCGCGGAGCGCATGACATCCGC 516
| | | | |
QY 541 GTCAACTGCTCAACNCGGAGCACTGCGCNGTCCGCTCGTGTCTTGAGGCGCTCAC 600
| | | | |
Db 517 GTCAACAGTTGACCCGCGGCACTCGGCTGTCCGCTCGTGTCTTGAGGCGCTCAC 576
| | | | |
QY 601 AACTTCCTCAACAGGATGATCAACCCCATGCTCCCGCTCGGCGGACATCTCGGCGTGG 660
| | | | |
Db 577 AACTTCCTCAACAGGATGATCAACCCCATGCTCCCGCTCGGCGGACATCTCGGCGTGG 636
| | | | |
QY 661 GCGGACTCTCCCGCCTCTCNTACATGCGCGCGCATCAACCGGTCAACCGGACNNAG 720
| | | | |
Db 637 GCGGACT 696
| | | | |
QY 721 GTNCAAGTNNTCAGAGAGGCGANNAGAAATCATGTNCGCCCGCGAGGATGCGCTC 780
| | | | |
Db 697 GTGCAAGTGTTCAGAGAGGCGAAGAGAAATCTGTAGCGCCCGAGGCGATGCGCTC 756
| | | | |
QY 781 TTNGCTTCGAGCCCGTGTCTCTCGGCCGAGAGAGGTCTCGTCTCGTCAACGCGACG 840
| | | | |
Db 757 TTCAACTTCGAGCCCGTGTCTCTCGGCCGAGAGAGGTCTCGTCTCGTCAACGCGAC 816
| | | | |
QY 841 GCGGTCTCGGCTGATGAGGCACTCTGCTTGCAAGAGCAACATGCTCTCGCTCTC 900
| | | | |
Db 817 GCGGTCTGAGCATGATGAGCACTCTCGCTTGCAAGAGCAACATGCTCTCGCTCTC 876
| | | | |
QY 901 TCGGAGGCGCTCAAGGCGCTTNAAGGTGAGGCGCATGTGTGCGGCGACGCGGCTCGTTCCAC 960
| | | | |
Db 877 TCGAGTGTCTCAAGGCGCATGAGGTGAGAGGATGTGTGCGGCGACGCGGCTCGTTCCAC 936
| | | | |
QY 961 CANTTCTTCAAGAGTCAAGCGCCCTTCAACCGCAACCGATGAGGTGCGGCAATC 1020
| | | | |
Db 937 CCTTCTCTTCAAGAGTCAAGCGGCGCTTCAACCGCAACCGATGAGGTGCGGCAATC 996
| | | | |
QY 1021 GCGAGCTCTTCAAGAGGCGAGGCGTTCGCTTCAACCGCAAGAGGAGGTGAAGTCAAG 1080
| | | | |
Db 997 GCGAGCTCTTCAAGAGGCGAGGCGTTCGCTTCAACCGCAAGAGGAGGTGAAGTCAAG 1056
| | | | |
QY 1081 GAGGAGGAGGATTTCTCGGCGAGGAGCGCTACCGGCTCCGCAAGTGAAGTGAAGTGC 1140
| | | | |
Db 1057 GAGGAGGAGGATTTCTCGGCGAGGAGCGCTACCGGCTCCGCAAGTGAAGTGAAGTGC 1116
| | | | |
QY 1141 GGCCTGCTGTGAGGAGATGATTTCAAGCCCAAGGNTCTCTGTGCTGAGGCGGNCAG 1200
| | | | |
Db 1117 GGCCTGCTGTGAGGAGATGATTTCAAGCCCAAGGNTCTCTGTGCTGAGGCGGNCAG 1176
| | | | |
QY 1201 TCGAGGAGCGCAACCGCGTATGAGGTGAGGAAACAAGANAGCCACAGCGGCGCAAC 1260
| | | | |
Db 1177 TCGAGGAGCGCAACCGCTTCTATGAGGTGAGGAAACAAGTTCGACCGGCGGCAAT 1236
| | | | |
QY 1261 TTTCAGGCGNCGCTGTGCAACCAAGATGAGAAAGACTGCTTCGCTCGCTGATC 1320
| | | | |
Db 1237 TTTCAGGCGTGGCTGTGCAACCAAGATGAGAAAGACTGCTTCGCTCGCTGATC 1296
| | | | |
QY 1321 GCGAGGCTAACTTCAAGAGAGTCAACCGAGATGCTCAACCGCGGATGAAACGCGGCTTN 1380
| | | | |
Db 1297 GCGAGGCTAACTTCAAGAGAGTCAACCGAGATGCTCAACCGCGGATGAAACGCGGCTTN 1356
| | | | |
QY 1381 CANTCTGCTGCTGCTGCGAGGAGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440
| | | | |
Db 1357 CCTCTGCTGCTGCGGCGGAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1416
| | | | |
QY 1441 GCGCGNCGNCTTACACTTGGAGCTCGAGNCACTTGGCAACCGGATTAACCAACGATC 1500
| | | | |
Db 1417 GCGCGNCGNCTTACACTTGGAGCTCGAGNCACTTGGCAACCGGATTAACCAACGATC 1476
| | | | |

QY 601 AACTTCCTCAACACGAGGATCAACCCCATCGTCCCTTCGCGGACCATCTCGAGCTCG 660
DB 577 AACTTCTCAACACGAGGATCAACCCCATCGTCCCTTCGCGGACCATCTCGAGCTCG 636
QY 661 GAGGACCTCTCCCT 720
DB 637 GAGGACCT 696
QY 721 GTTACAGTNNNNNAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780
DB 697 GTTACAGTNNNNNAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 756
QY 781 TTNGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
DB 757 TTACACTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 816
QY 841 GCGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 900
DB 817 GCGCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 876
QY 901 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 960
DB 877 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 936
QY 961 CCGTCTCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
DB 937 CCGTCTCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996
QY 1021 CGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 997 CGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1056
QY 1081 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1140
DB 1057 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1116
QY 1141 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1200
DB 1117 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1176
QY 1201 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1260
DB 1177 TCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1236
QY 1261 TTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
DB 1237 TTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1296
QY 1321 GCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1380
DB 1297 GCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1356
QY 1381 CCGTCTCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
DB 1357 CCGTCTCTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1416
QY 1441 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
DB 1417 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1476
QY 1501 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1560
DB 1477 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1536
QY 1561 ACNCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
DB 1537 ACNCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1596
QY 1621 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1680
DB 1597 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1656

QY 1681 NCNCGCTCNTCMAGCAGCACTTGGGCGNCGCCCTCGAGGCGNNNNNNNNNNNNNN 1740
DB 1657 GTTCTGCTCATCGACGAGCACTTGGGCTCGCCATGACCGGCTCGAACTCGCCGAGG 1716
QY 1741 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
DB 1717 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1776
QY 1801 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1860
DB 1777 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1836
QY 1861 NNGTCTCTCCNNNNCCANNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
DB 1837 TCGTGAAGGTC-----GCTCTGCTCGCGCGCGGAGGAGGAGGAGGAGGAGG 1885
QY 1921 CGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
DB 1886 CGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1945
QY 1981 CCGTCTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
DB 1946 CGACTCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2005
QY 2041 TCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100
DB 2006 TCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2065
QY 2101 TGACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2160
DB 2066 TGACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2125
QY 2161 ACGTCTCTCAAGATGCTCGCCTAG 2186
DB 2126 ACGTCTCTCAAGATGCTCGCCTAG 2151

RESULT 13
US-10-188-523B-7
; Sequence 7, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; PRIOR FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2151)
; OTHER INFORMATION:
US-10-188-523B-7

Query Match 65.6%; Score 1623.2; DB 15; Length 2151;
Best Local Similarity 83.6%; Pred. No. 0;
Matches 1827; Conservative 0; Mismatches 324; Indels 35; Gaps 5;
QY 1 ATGCCCCCTCCTGAGTCTGATGCGAAGCTCGTGCCTCAAGGAGGAGGAGGAGGAGG 60
DB 1 ATGAGCACTCTGCTGATCTGATCTGCACTCTGCTGCAAGCGAGGAGGAGGAGGAGG 60
QY 61 CAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120

Db 61 CAGGCTGT-----CAATGGCGCTTGACCAACCTC-----GCACTCCGAG 100
 QY 121 GCTGCTCTCTCCGACCAACCCGANNNGACGAGCTGACATCTGTNGAGNAGATCTCTCCGC 180
 Db 101 GCTCGACCTTGCCCAACCCGAGGTCAAGCAGGTGACATGCTGAGAGAGTGTCTCCGCG 160
 QY 181 ACCCCACCGNNMAGNACGNNMTCGAACGTGACGGGTACACCTTACCTCGNNACGTG 240
 Db 161 CGCCGACCACTGCA---CGCTGAACTGAGGGCTACTCTCAACTCGAGACGTG 217
 QY 241 TCGGCGCCGNCGCAAGGCGCCGNCNGTCCGCTGNCAGACAGNCGACGAGATCCGCGCA 300
 Db 218 TCTCGGCGCGGAGGAGGCAAGCGCTGTCCGCTCAAGAGACAG--GAGAGNATCCGCTCA 276
 QY 301 AAGATGCAAAAGATGAGTGTCTCTCCGNNCAGCTGACCAACAGNCTAGAGNCTC 360
 Db 277 AAGATTGAAATACTCGGTGAGTCTTGGCTGCGCACTCTCATGAGGCTGTACGGGCTC 336
 QY 361 ACGACTGTTTTGCGCGGCTCGGCGGACACCGGACTGAGATGATGATCTGCTCAGAG 420
 Db 337 ACGACTGATTTGGGGGATCCGACAGACCCGACCGAGAGCGCATCTCGCTCAAGAG 396
 QY 421 GGNCTCTCGAGACCAAGCTCTGCGGTGTCTCCGACGTCGANTCGANTCTTCCGCTC 480
 Db 397 GCTCTCTCGACACACAGCTCTGCGGTGTCTCTCTGCTGCTGCACTCGTTCCGCTC 456
 QY 481 GGNCGGCGCTCGAGAACTGCTTCCGCTCGAGGTGCTGCGGGCGGACATGACATCGC 540
 Db 457 GCGCGGCTCTGAGAACTCGCTTCCCTCGAGGTGTTTGGGGCGGACATGACATCGC 516
 QY 541 GTCAACTCGCTACNCGCGGCACTCGGNGTCCGCTCTGCTCTCGAGGCGCTCAC 600
 Db 517 GTCAAGAGTTGACCCGCGGCGCACTGGCTGTCCGCTGTCTGCTCGAGGGCTCAC 576
 QY 601 AACTTCTCTAACAAGGATCAACCCCATCTGCTCTCCGCGGACCATCTCGGCTCG 660
 Db 577 AACTTCTCTAACAAGGATCAACCCCATCTGCTCTCCGCGGACCATCTCGGCTCG 636
 QY 661 GCGGACCTCTCCGNCCTGTCATGATCGCGCGGCACTACCGGTCACCGGAGCANNAG 720
 Db 637 GCGGACCT 686
 QY 721 GTNCACTGNNNTACGAGGGGACNNAGAAATCATGTNCGCGCGGAGCGATCGCTC 780
 Db 697 GTGCACTGCTGACGAGGGGACGAGAAATCTCTGACCGCGGAGCGATGCGCTC 756
 QY 781 TTAGGCTCGAGCCGCTGCTCTGCGCGGAGAGAGGTCTCTGCTCTCTCAACGGACG 840
 Db 757 TTCAACCTGAGCCCGTCTCTGCGCGGAGAGAGGTCTCTGCTCTCTCAACGGACG 816
 QY 841 GCGCTCTCGGCTCGATGAGGACCCCTGCTCTGACGACGACACATGCTCTGCTCTC 900
 Db 817 GCGCTCTGACATGATGAGGACCCCTGCTCTGACGACGCTCAATGCTCTGCTCTC 876
 QY 901 TCGCAGAGGCTCAGGCTCTTACGATGAGGACATGCTGCGGCGGCTCTCTCTCA 960
 Db 877 TCGCAGTGTCTACGCGCATGACGATGAGGATGCTGCGGCGGCGGCTCTCTCTCA 936
 QY 961 CANTTCTCTCAGAGATCAAGGCGCTTACCGGACCGGATGAGATGAGGCTGCGGCAATC 1020
 Db 937 CCTTCTCTTCAAGAGATCAAGGCGCTTACCGGACCGGATGAGATGAGGCTGCGGCAATC 996
 QY 1021 CGCAGCTCTCTGAGGAGCAGNNGTGCGCTCAACGAGAGAGAGTCAAGTCAAG 1080
 Db 997 CCGAAGCTCTGAGAGGAGCGCTTCTCTCTCAACGAGAGAGAGTCAAGTCAAG 1056
 QY 1081 GAGCAGAGAGGATCTCTCGCAGAGACGCTACCGGCTCGGACGCTCTGAGTCTC 1140
 Db 1057 GAGCAGAGAGGATCTCTCGCAGAGACGCTACCGGCTCGGACGCTCTGAGTCTC 1116
 QY 1141 GCGCGCTCTCTGAGGAGATGATTCACGCGCACGCGNCTCTCTGCTGAGGCGGAG 1200
 Db 1117 GCGCGCTCTCTGAGGAGATGATTCACGCGCACGCGNCTCTCTGCTGAGGCGGAG 1176

QY 1201 TCGACGACGACCAACCGCTCATGAGCTGAGAGCAAGANAGACCAACGAGGCAAC 1260
 Db 1177 TCGACGACGACCAACCGCTCATGAGCTGAGAGCAAGATTTGACACGAGGCAAT 1236
 QY 1261 TTCGAGGCGCGCTGTGCAACAGATGAGAGAGCTGCGCTGCGNCTGCGCTGATC 1320
 Db 1237 TTCGAGGCTGCGCTGTGCAACAGATGAGAGAGCTGCGCTGCGNCTGCGCTGATC 1296
 QY 1321 GCGAAGCTCAATTCAACGAGCTCAACGAGATGCTCAACGCGGATGAACCGGCTTN 1380
 Db 1297 GCGAAGCTCAATTCAACGAGCTCAACGAGATGCTCAACGCGGATGAACCGGCTTN 1356
 QY 1381 CCGTCTGCTGCTGCGGAGAGACCCCTGCTCTCTATCACTGCAAGGCTTGAAT 1440
 Db 1357 CCGTCTGCTGCGGAGAGACCCCTGCTCTCTATCACTGCAAGGCTTGAAT 1416
 QY 1441 GCGGNGGNCNTACTTGGAGCTCGANACTTCCCAACCGGCTTNAAGACCAAGTC 1500
 Db 1417 GCGGCTGCGGCTGACCTCGGAGTGGAGACCTCGGCAACCTGTGAGACGCAATC 1476
 QY 1501 CAGCGGCGNAGATGAGCAACGAGCGCTCAACTGCTGCGNCTCATCTGCGNCGCGC 1560
 Db 1477 CAGCGGCTGAGATGAGCAACGAGCGCTCAACTGCTGCGNCTCATCTGCGNCGCGC 1536
 QY 1561 ACMGCGAGGCGCAACGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
 Db 1537 ACMGCGAGTCAACGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1596
 QY 1621 CAGGCGCTGCACTCTCGGCGATGAGATGAGTCAAGTCAAGAGAGTCAACCGGNTNTC 1680
 Db 1597 CAGGCGCTGCACTCTCGGCGATGAGATGAGTCAAGTCAAGAGAGTCAACCGGNTNTC 1656
 QY 1681 NNNNGGCTCTNACGAGACCTTTGGGCGNCGCTCGACGCGGNNNNNNNNNNNAGAA 1740
 Db 1657 GTCTGCTCATGACGAGACATTTGGCTCGGCAATGACGCGGCTGAACTCGCGGAG 1716
 QY 1741 CTCGNGGCAAGAGTCAACGAGCTGNAACGAGCTCGAGGACCAACTGTAAGAC 1800
 Db 1717 CTCGNGGAGAGTCAACGAGCTGNAACGAGCTGNAACGAGCTGTAAGAC 1776
 QY 1801 CTCGAGCGGCTGAGCAGACGCTTCTGTTGCGGACCGGACCGCTGTAAGNNNNN 1860
 Db 1777 CTCGAGCGGCTGAGCAGACGCTTCTGTTGCGGACCGGACCGCTGTAAGNNNNN 1836
 QY 1861 NNGTCTGCGNNNNNGCANNAGGCTGCTGCGNCGGCTCAACGCTGAAAGGTGCT 1920
 Db 1837 TCGTGAAGTC-----GCTTGGCTGCGGCGGCTGTAAGGCTGCGG 1885
 QY 1921 CCGCGGAGAGGCACTCTGCTCAACGCGNAGTCCGCAACNCTTCTGAGCGGCGT 1980
 Db 1886 CCGCGGAGTCCGCGCATCTGCTCAACGCGGCAAGTCCGCGGAGCTTGTGCGGCGGT 1945
 QY 1981 CGTGTCTGCGCGCGGCTGCGGATCTTCTGCGGCGGAGCGGCTTGTATCTGTTG 2040
 Db 1946 CGACTGTGCGCGGCTGCGGATCTTCTGCGGCGGAGCTTGTATCTGTTG 2005
 QY 2041 TCGCGGAGAGCTGCGGCTCAAGGCGCGCGGCGGAGCTTCTTCTGCGGAGGAGG 2100
 Db 2006 TCGCGGAGAGCTTGGCTCAAGGCGCGCGGCGGAGAGCTTCTTCTGCGGAGGAGG 2065
 QY 2101 TCGAGATGAGCAACGCTCTCCGCACTTACGAGGCTCAAGGCGGCGGATCAAC 2160
 Db 2066 TCGAGATGAGCTCAACGCTCTCCGCACTTACGAGGCTCAAGGCGGCGGATCAAC 2125
 QY 2161 AGTCTCTGTAAGTGTGCGNTAG 2186
 Db 2126 AGTCTCTGTAAGTGTGCGNTAG 2151

RESULT 14
 us-10-374-366-1
 ; Sequence 1, Application US/10374366

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:01:03 ; Search time 23.5975 Seconds
(without alignments)
1598.322 Million cell updates/sec

Title: US-09-939-408A-21
Perfect score: 3365
Sequence: 1 MAPSDISATXSXANGXNGX.....RIYAIKGRINXVYVLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 369414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3223	95.8	726	4	US-09-624-693A-21
2	3021	89.8	716	4	US-09-624-693A-19
3	3021	89.8	716	4	US-09-627-216A-8
4	3021	89.8	716	4	US-09-765-873A-8
5	3019	89.7	716	4	US-09-627-216A-10
6	3019	89.7	716	4	US-09-765-873A-10
7	3014	89.6	716	4	US-09-765-873A-38
8	3012	89.5	716	4	US-09-765-873A-35
9	3010	89.5	716	4	US-09-765-873A-36
10	3010	89.5	716	4	US-09-765-873A-37
11	3006	89.3	716	4	US-09-765-873A-34
12	2978	88.5	716	4	US-09-765-873A-33
13	2971	88.3	686	4	US-09-765-873A-32
14	2922.5	86.8	713	4	US-09-624-693A-17
15	2770	82.3	720	4	US-09-624-693A-13
16	1294.5	38.5	740	4	US-09-624-693A-15
17	670	19.9	424	4	US-09-615-192A-128
18	495.5	14.7	515	4	US-09-328-352-5090
19	490	14.6	508	4	US-09-489-039A-13575
20	480.5	14.3	515	4	US-09-252-991A-13575
21	474.5	14.1	525	4	US-09-540-236-2855
22	459	13.6	412	4	US-09-615-192A-130
23	365	10.8	520	4	US-09-252-991A-13049
24	223	6.9	226	4	US-09-615-192A-327
25	223.5	6.6	97	4	US-09-615-192A-329
26	115	3.4	1024	4	US-09-562-737-50
27	111	3.3	1024	4	US-09-562-737-41

28	108	3.2	1024	4	US-09-562-737-44	Sequence 44, App1
29	106	3.2	505	4	US-09-252-991A-23925	Sequence 23925, A
30	104	3.1	744	1	US-08-764-100-25	Sequence 25, App1
31	102.5	3.0	1011	4	US-09-252-991A-22505	Sequence 22505, A
32	100.5	3.0	555	4	US-09-543-681A-7878	Sequence 7878, Ap
33	100	3.0	1589	4	US-09-543-681A-4958	Sequence 4958, Ap
34	99.5	3.0	1024	4	US-09-562-737-45	Sequence 45, App1
35	96.5	2.9	404	4	US-09-328-352-6480	Sequence 6480, Ap
36	96	2.9	875	4	US-09-328-352-4884	Sequence 4884, Ap
37	95.5	2.8	836	4	US-09-252-991A-23513	Sequence 23513, A
38	94.5	2.8	607	3	US-08-969-683A-59	Sequence 59, App1
39	94	2.8	555	1	US-08-440-377A-6	Sequence 6, App1
40	94	2.8	555	2	US-08-687-852-6	Sequence 6, App1
41	94	2.8	764	4	US-09-252-991A-31816	Sequence 31816, A
42	93.5	2.8	1024	4	US-09-562-737-47	Sequence 47, App1
43	93.5	2.8	1024	4	US-09-562-737-49	Sequence 49, App1
44	93.5	2.8	1495	4	US-08-522-726B-1	Sequence 1, App1
45	93.5	2.8	1495	4	US-09-337-384-1	Sequence 1, App1

ALIGNMENTS

RESULT 1
US-09-624-693A-21
; Sequence 21, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koestra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; NAME/KEY: SITE
; LOCATION: (12, 16, 17, 20, 24, 25, 28 - 36, 38, 42, 47, 48, 56, 57, 62, 66,
; OTHER INFORMATION: "Xaa" means any amino acid; "Xaa" means no consensus at that
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
; OTHER INFORMATION: SEQ ID NOS: 13, 17, and 19
US-09-624-693A-21

Query Match 95.8%; Score 3223; DB 4; Length 726;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSDISATXSXANGXNGXAAAXAXXXXXXXXAGSXPETXTTOLDIVEXXLD 60
1 MAPSDISATXSXANGXNGXAAAXAXXXXXXXXAGSXPETXTTOLDIVEXXLD 60
Db 1 MAPSDISATXSXANGXNGXAAAXAXXXXXXXXAGSXPETXTTOLDIVEXXLD 60
QY 61 PXTDDXHEIDGYSLTLDGVGAARXVXVDSDEIRXKIDKSVBELRXQLXNSVYGV 120
61 PXTDDXHEIDGYSLTLDGVGAARXVXVDSDEIRXKIDKSVBELRXQLXNSVYGV 120
Db 61 PXTDDXHEIDGYSLTLDGVGAARXVXVDSDEIRXKIDKSVBELRXQLXNSVYGV 120
QY 121 TFGGSAATREDASISOKALLEHOLCGVLTSDXSYLXGLGNSLPLEVVRGAMTIRV 180
121 TFGGSAATREDASISOKALLEHOLCGVLTSDXSYLXGLGNSLPLEVVRGAMTIRV 180
Db 121 TFGGSAATREDASISOKALLEHOLCGVLTSDXSYLXGLGNSLPLEVVRGAMTIRV 180
QY 181 NSLTGSHAVRLVLEALTNFLNIGTPIVPLRGTSASGDLSPSTYIAAITHGPSKV 240
181 NSLTGSHAVRLVLEALTNFLNIGTPIVPLRGTSASGDLSPSTYIAAITHGPSKV 240
Db 181 NSLTGSHAVRLVLEALTNFLNIGTPIVPLRGTSASGDLSPSTYIAAITHGPSKV 240
QY 241 HXHEGXKIXKXARBAALFGLPEVYVLPKGLGVNGTAVASAMATLALDAMTSLLS 300
241 HXHEGXKIXKXARBAALFGLPEVYVLPKGLGVNGTAVASAMATLALDAMTSLLS 300

```

Dh      241  HVXHEXKEMAREALIALFGJEPVYLGPKESTGLVNGTAVASASVATLALHDAMLSLS 3000
Qy      301  QALTALTVEAMVGHAGSFHPEPLHDVTRHPPTQIEVARNIRTLBGSXAVNHEEYVXD 3600
Dh      301  QALTALTVEAMVGHAGSFHPEPLHDVTRHPPTQIEVARNIRTLBGSXAVNHEEYVXD 3600
Qy      361  DEGLIKODRPLATSPQWLGPLVSDMHAHAAVLSLEAGOSTTNDPLIDVENKXTTHGANE 4200
Dh      361  DEGLIKODRPLATSPQWLGPLVSDMHAHAAVLSLEAGOSTTNDPLIDVENKXTTHGANE 4200
Qy      421  QASAVXNTMEKTRLTALALIGKLNFTOLTETMLNAGNNGLPSCLAEDPSSLTHCKGLDIA 4800
Dh      421  QASAVXNTMEKTRLTALALIGKLNFTOLTETMLNAGNNGLPSCLAEDPSSLTHCKGLDIA 4800
Qy      481  AAAYTSELGLANPVTTHQOPABMGNOAVNSLALISARTABADVLSLLATLTYCLO 5400
Dh      481  AAAYTSELGLANPVTTHQOPABMGNOAVNSLALISARTABADVLSLLATLTYCLO 5400
Qy      541  AVDLRAMEPEFKQEPKXXXLIXHGFQXXTXXXXXELKXXYKXLKRLKQETNSYDL 6000
Dh      541  AVDLRAMEPEFKQEPKXXXLIXHGFQXXTXXXXXELKXXYKXLKRLKQETNSYDL 6000
Qy      601  EPRMHDAFSXATGTVEEXLSXXXXXVSLAAVANAKVAXAEKASLTRYVXXFXWAPSS 6600
Dh      601  EPRMHDAFSXATGTVEEXLSXXXXXVSLAAVANAKVAXAEKASLTRYVXXFXWAPSS 6600
Qy      661  SSPALKYLSPPRTVLYSFPVRELGVKARPGVFLGKQSVTIGSNVSRLYEAKSGRINXV 7200
Dh      661  SSPALKYLSPPRTVLYSFPVRELGVKARPGVFLGKQSVTIGSNVSRLYEAKSGRINXV 7200
Qy      721  LVKMLA 726
Dh      721  LVKMLA 726

RESULT 2
US-09-624-693A-19
/ Sequence 19, Application US/09624693A
/ Patent No. 6355468
/ GENERAL INFORMATION:
/ APPLICANT: Yoshida, Roberta
/ TITLE OF INVENTION: Koostera, Anna
/ TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
/ TITLE OF INVENTION: Polymaleotide Sequences and Methods of Obtaining and
/ FILE REFERENCE: 29479/500NEC
/ CURRENT APPLICATION NUMBER: US/09/624,693A
/ CURRENT FILING DATE: 2000-07-24
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 19
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Rhodococcus toruloides
/ US-09-624-693A-19

```

Query Match	89.8%	Score 3021	DB 4	Length 716
Best Local Similarity	84.0%	Pred. No. 0		
Matches 610	Conservative 25	Mismatches 81	Indels 10	Gaps 3
QY	1	MAPSLDLSIATSYANGXNGXHAAXXASXXXXXXXXXXAGSLPTTXXTQDIVEXXLAD	60	
Db	1	MAPSLDISISHSPANGVASAKQAVNGAS-----TNLVAASHLPTTQVQVDIVEKMLAA	54	
QY	61	PXIDDXAHELDGSLTLGDIVGAARGRGRYRXDSDEFKXIDKSVEFLRQXLSNXYGT	120	
Db	55	P-VDSTLELDGSLNMGDVSAARGRGPRYKDSDEFKXIDKSVEFLRQLSMSYGYT	113	
QY	121	TGFGGSAQDRTEDASLQKALLLEHQLCVLPTSDXSEFLRGLENSLPLEVVGAMTIRY	180	
Db	114	TGFGGSAQDRTEDASLQKALLLEHQLCVLPTSDPRLRGLENSLPLEVVGAMTIRY	173	
QY	181	NSITRGHSAVRLVLEALTNPLNHSITPIVLRGTTISAGDSPLSYIAAALTGHPDSKY	240	

```

Db      174 NSLIRHSANVLVLVEALTNFLNHSITITVLPLRGITISAGDLSPLSTIAAISGHPDSKV 233
Qy      241 HXHXEKGEXKIMAREEAIALFGLBRYVLPCKGGLVNGTANSASMATLALHDAMLSLS 300
Db      234 HVHBEKEKILVAREMALFNLBEPVLDPCKGLGVNGTASASMATLALHDAMLSLS 293
Qy      301 QALTAITVEMNGHAGSFHPLHDVTRDHPQIEVARNIRPTLLEGSXFEAVHHEEYKVD 360
Db      294 QSLTANTVEMNGHAGSFHPLHDVTRDHPQIEVAGNIRKLEGGSRFAVHHEEYKVD 353
Qy      361 DECILRQDSYPLRTSPQWIGPLVDSMIIHAHVLISLEAGOSTTNDPLIDVENKXKTHGQNF 420
Db      354 DEGLIRQDSYPLRTSPQWIGPLVDSMIIHAHVLITLEGQOSTTNDPLIDVENKXKTHGQNF 413
Qy      421 QASAVNNTKEXRLALALIGKLNFPQLHEMNAGMGNLPSCLAAEDPSLSYHCKGLDIA 480
Db      414 QASAVNNTKEXRLALALIGKLNFPQLHEMNAGMGNLPSCLAAEDPSLSYHCKGLDIA 473
Qy      481 AAAYTESELGLANPYTHVQPAEMGNQAVNSLALISARTEANDVLSLLATHLXYCULQ 540
Db      474 AAAYTESELGLANPYTHVQPAEMGNQAVNSLALISARTEANDVLSLLATHLXYCULQ 533
Qy      541 AVULRAMEREPFKQKQRPXXKXLLXHPKQKXTXXXXXXELXKTKXKXKXRLDQTSYDL 600
Db      534 AILRLRAIBEFKQKQRPALVSLIDHFSAMTGSNRLDELVEKNKTLARLEQTSYDL 593
Qy      601 EPRMHDAPFXATGTVEXLSKXXXXVSLAANMKAAXAEKALSTRXYRXKFWXAPS 660
Db      594 VPRMHDAPFXATGTVEXLSKXXXXVSLAANMKAAXAEKALSTRXYRXKFWXAPS 650
Qy      661 SSPALXYLSPRTVLYSFVREELGYKARQGVFGKQEVITGSNVSKITYAIGSRINXY 720
Db      651 SSPALXYLSPRTVLYSFVREELGYKARQGVFGKQEVITGSNVSKITYAIGSRINXY 710
Qy      721 LVYKOLA 726
Db      711 LVYKOLA 716

```

```

RESULT 3
US-09-627-216A-8
; Sequence 8, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; PRIOR FILING DATE: 2000-07-27
; PRIOR FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-09-627-216A-8

Query Match      89.8%; Score 3021, DB 4; Length 716;
Best Local Similarity 84.0%; Freq. No. 0;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3

Oy      1 MAPSLDIITSXANGXNGXHAAXXASXXXXXXAGSXLPPTXXTQLDIVEXXLD 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MAPSLDISHSANGVASAKQAVNGAS-----TNLAVASHPPTQTQVQVDIVELGLAA 54
Oy      61 PXTDDXELLEDGYSILGDDVGAARKGRKVRXVDSDFRXXIDKSVFLKXQLXNSVYGT 120

```


Db 55 P-TDSTLELDGSLNIGDVASARKGRPVXKDSDEIRSKIDKSVFRLSGLNSVYGV 113
QY 121 TGFSGSADTRTEDAISLQKALLEHQLCVLPSTXDSFXLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCVLPSTXDSFXLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGHSARLVYLVLEALTNFNLHGITPIVPLRGITISASGDISPLSYAAAITGHPDSKV 240
Db 174 NSLTRGHSARLVYLVLEALTNFNLHGITPIVPLRGITISASGDISPLSYAAAITGHPDSKV 233
QY 244 HYHGXEXKIMKAREALFGLPEVVLGPKGGLGVNGTAVSASMATLALDAMLSLIS 300
Db 234 HYHGXEXKILYAREMALLFNLEPVVLGPKGGLGVNGTAVSASMATLALDAMLSLIS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSXFAYHHEEVKVD 360
Db 294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGSRFAYHHEEVKVD 353
QY 361 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGDNF 420
Db 354 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGDNF 413
QY 421 QASAVANTMEKTRIALALIGKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRIRLQAGIKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATLHYCYLQ 540
Db 474 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAESNDVLSLLATLHYCYLQ 533
QY 541 AAYDLRAMEEFKQFPPKQFXXXXLLXOHFGXXXXXXELKXXKXKXRELEGTNSYDL 600
Db 534 AAYDLRAMEEFKQFPPKQFPAIVSLIDHFGSANTGSLRDELVEKNKTLAKLEGTNSYDL 593
QY 601 EPRMHDAPFSGATGVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXFXMXPSS 660
Db 594 VPRMHDAPFSGATGVVEVLSS--TSLSLAAVNAKVAAXEKAISLTRQVETFWMSAST 650
QY 661 SSPALXYLSPPRTVLXSYFVREBELGVKARRGDVFLLGKQVTTGNSVSKRYEAIKSGRINXV 720
Db 651 SSPALXYLSPPRTQILYAFVREBELGVKARRGDVFLLGKQVTTGNSVSKRYEAIKSGRINXV 710
QY 721 LVKMLA 726
Db 711 LVKMLA 716

RESULT 4

US-09-765-873A-8
; Sequence 8, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BCI009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIORITY FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 09/627, 216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147, 719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-09-765-873A-8

Query Match 89.8%; Score 3021; DB 4; Length 716;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 610; Conservative 25; Mismatches 91; Indels 10; Gaps 3;

QY 1 MAPSLDSTATSANGXNGXHAAXXXXXXXXAGSXLPTTXXTQOLDIVEXKLAD 60
Db 1 MAPSLDSTATSANGXNGXHAAXXXXXXXXAGSXLPTTXXTQOLDIVEXKLAD 54
QY 61 PXTDDXHELDDGSLTLGADVGAARKGRVXKDSDEIRSKIDKSVFRLSGLNSVYGV 120
Db 55 P-TDSTLELDGSLNIGDVASARKGRPVXKDSDEIRSKIDKSVFRLSGLNSVYGV 113
QY 121 TGFSGSADTRTEDAISLQKALLEHQLCVLPSTXDSFXLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCVLPSTXDSFXLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGHSARLVYLVLEALTNFNLHGITPIVPLRGITISASGDISPLSYAAAITGHPDSKV 240
Db 174 NSLTRGHSARLVYLVLEALTNFNLHGITPIVPLRGITISASGDISPLSYAAAITGHPDSKV 233
QY 244 HYHGXEXKIMKAREALFGLPEVVLGPKGGLGVNGTAVSASMATLALDAMLSLIS 300
Db 234 HYHGXEXKILYAREMALLFNLEPVVLGPKGGLGVNGTAVSASMATLALDAMLSLIS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSXFAYHHEEVKVD 360
Db 294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLEGSRFAYHHEEVKVD 353
QY 361 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGDNF 420
Db 354 DEGITLRDORYPARTSPQWLGPIVSDMIAHAVALTIEAGOSTTDNPLIDVENKTSHHGDNF 413
QY 421 QASAVANTMEKTRIALALIGKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRIRLQAGIKLNFTQITEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATLHYCYLQ 540
Db 474 AAAYTSELGHLANPVTTHVQPAEMGNQAVNSLALISARRTAESNDVLSLLATLHYCYLQ 533
QY 541 AAYDLRAMEEFKQFPPKQFXXXXLLXOHFGXXXXXXELKXXKXKXRELEGTNSYDL 600
Db 534 AAYDLRAMEEFKQFPPKQFPAIVSLIDHFGSANTGSLRDELVEKNKTLAKLEGTNSYDL 593
QY 601 EPRMHDAPFSGATGVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXFXMXPSS 660
Db 594 VPRMHDAPFSGATGVVEVLSS--TSLSLAAVNAKVAAXEKAISLTRQVETFWMSAST 650
QY 661 SSPALXYLSPPRTVLXSYFVREBELGVKARRGDVFLLGKQVTTGNSVSKRYEAIKSGRINXV 720
Db 651 SSPALXYLSPPRTQILYAFVREBELGVKARRGDVFLLGKQVTTGNSVSKRYEAIKSGRINXV 710
QY 721 LVKMLA 726
Db 711 LVKMLA 716

RESULT 5

US-09-627-216A-10
; Sequence 10, Application US/09627216A
; Patent No. 636837
; GENERAL INFORMATION:
; APPLICANT: Satiaslani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BCI009 US NA
; CURRENT APPLICATION NUMBER: US/09/627, 216A
; PRIORITY FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147, 719
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10

LENGTH: 716
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: mutant from
US-09-627-216A-10

Query Match 89.7%; Score 3019; DB 4; Length 716;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

QY 1 MAFSLDSTATSANGKXNKAAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXIAD 60
DB 1 MAFSLDSTATSANGKXNKAAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXIAD 54
QY 61 PXTDDXXELDGYSLTIGDVGGAARKGRVRYXDSDEIRKIDKSYVEFLRXQLXNSVYGV 120
DB 55 P-TDSTELDGYSLNIGDVSAARKGRPRVXKDSDEIRKIDKSYVEFLRQLSMSVYGV 113
QY 121 TGFSGADRTEDALSLQKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 180
DB 114 TGFSGADRTEDALSLQKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISAGDLSPLSYIAAIGHPDSKV 240
DB 174 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISAGDLSPLSYIAAIGHPDSKV 233
QY 241 HVXHEGKEKIMXAREALFGLPEVVLGPKGGLVNGTAVSASMATLALDAMLSLS 300
DB 234 HVXHEGKEKILYAREAMALFNLPEVVLGPKGGLVNGTAVSASMATLALDAMLSLS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSAFAVHEBEVVKD 360
DB 294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSAFAVHEBEVVKD 353
QY 361 DGLIRQDRYPLRTSPQWGLPVSDMIHAVALSLAAGOSTTDNPLIDVENKXTHHGNF 420
DB 354 DGLIRQDRYPLRTSPQWGLPVSDMIHAVALSLAAGOSTTDNPLIDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGMNGRLPSCLAEDPSLSYHCKGLDIA 480
DB 414 QAAAVANTMEKTRIALALIGKLNFTQLTMLNAGMNGRLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPVTHVQPAEMGQAVNSLALISARTAEANDVLSLLATHLVCLQ 540
DB 474 AAAYTSELGHLANPVTHVQPAEMGQAVNSLALISARTTESNDVLSLLATHLVCLQ 533
QY 541 AVDLRAHEFEFKQFPYXXXLLXQHFQXXXTXXXXXXELXXXKXKXLEKLEQNSYDL 600
DB 534 AIDLRATEFEFKQFQPAIVSLIDQHFQSAMTGSNRLDELVEKNKTLAKLEQNSYDL 593
QY 601 EPRMHDASXATGTVEVLSXXXXXVSLAAVNAKVAEKAISLITXVXXFXKAPSS 660
DB 594 VPRMHDASFAAGTVEVLS--TSLSLAAVNAKVAEKAISLITRQVETFEFSAAST 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGQEVLTIGSNVSRITYEAIKSGRINXV 720
DB 651 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGQEVLTIGSNVSRITYEAIKSGRINNV 710
QY 721 LVXKMLA 726
DB 711 LKXKMLA 716

RESULT 6
US-09-765-873A-10
Sequence 10, Application US/09765873A
Patent No. 6521748
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOREDUCTION OF PABA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BC1009 US CIP

CURRENT APPLICATION NUMBER: US/09/765, 873A
CURRENT FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: US 09/627, 216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147, 719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 10
LENGTH: 716
TYPE: PRT
ORGANISM: mutant from Rhodococcus glutinis
US-09-765-873A-10

Query Match 89.7%; Score 3019; DB 4; Length 716;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

QY 1 MAFSLDSTATSANGKXNKAAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXIAD 60
DB 1 MAFSLDSTATSANGKXNKAAXXXXXXXXXXXXXXSLPTTXXXTQLDIYEXXIAD 54
QY 61 PXTDDXXELDGYSLTIGDVGGAARKGRVRYXDSDEIRKIDKSYVEFLRXQLXNSVYGV 120
DB 55 P-TDSTELDGYSLNIGDVSAARKGRPRVXKDSDEIRKIDKSYVEFLRQLSMSVYGV 113
QY 121 TGFSGADRTEDALSLQKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 180
DB 114 TGFSGADRTEDALSLQKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISAGDLSPLSYIAAIGHPDSKV 240
DB 174 NSLTRGSAVRLVLEALTNFLNHSITPIVPLRGITISAGDLSPLSYIAAIGHPDSKV 233
QY 241 HVXHEGKEKIMXAREALFGLPEVVLGPKGGLVNGTAVSASMATLALDAMLSLS 300
DB 234 HVXHEGKEKILYAREAMALFNLPEVVLGPKGGLVNGTAVSASMATLALDAMLSLS 293
QY 301 QALTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSAFAVHEBEVVKD 360
DB 294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSAFAVHEBEVVKD 353
QY 361 DGLIRQDRYPLRTSPQWGLPVSDMIHAVALSLAAGOSTTDNPLIDVENKXTHHGNF 420
DB 354 DGLIRQDRYPLRTSPQWGLPVSDMIHAVALSLAAGOSTTDNPLIDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGMNGRLPSCLAEDPSLSYHCKGLDIA 480
DB 414 QAAAVANTMEKTRIALALIGKLNFTQLTMLNAGMNGRLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLANPVTHVQPAEMGQAVNSLALISARTAEANDVLSLLATHLVCLQ 540
DB 474 AAAYTSELGHLANPVTHVQPAEMGQAVNSLALISARTTESNDVLSLLATHLVCLQ 533
QY 541 AVDLRAHEFEFKQFPYXXXLLXQHFQXXXTXXXXXXELXXXKXKXLEKLEQNSYDL 600
DB 534 AIDLRATEFEFKQFQPAIVSLIDQHFQSAMTGSNRLDELVEKNKTLAKLEQNSYDL 593
QY 601 EPRMHDASXATGTVEVLSXXXXXVSLAAVNAKVAEKAISLITXVXXFXKAPSS 660
DB 594 VPRMHDASFAAGTVEVLS--TSLSLAAVNAKVAEKAISLITRQVETFEFSAAST 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGQEVLTIGSNVSRITYEAIKSGRINXV 720
DB 651 SSPALXYLSPTRTVLVSFVREELGVKARGDVFLGQEVLTIGSNVSRITYEAIKSGRINNV 710
QY 721 LVXKMLA 726
DB 711 LKXKMLA 716

RESULT 7
US-09-765-873A-38

```

; Sequence 38, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 38
; LENGTH: 716
; TYPE: PRP
; ORGANISM: mutant from Rhodocorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: X = Gly, Ala, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X = Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

Query Match      89.6%; Score 3014; DB 4; Length 716;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

QY      1 MAPSLDSTISXANGXNKGKHAAXAAXXXXXXXXAGXSLTTYXTQLDIYEXXLD 60
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1 MAPSLDSTISHSFANGVSAKQAVNGAS-----TNLAAGSHLPTTYQVDIYEXKMLAA 54
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      61 PXTDXXELDGYSLTLGQVVGAAKRGKXVRYXDSDEIRXKIDKSVFPRQIXNSVYGT 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      55 P-TDSTLELDGYSNLGQVVSAAKRGKRPVRYKDSDEIRSKIDKSVFPRQIXNSVYGT 113
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      121 TGFSGSADTRTEDAISLOKALLEHQLCGVLPFSXDSFXLGRLENSLPLEVVGAMTIRV 180
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      114 TGFSGSADTRTEDAISLOKALLEHQLCGVLPFSXDSFXLGRLENSLPLEVVGAMTIRV 173
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      181 NSLTRGHSNVRVLVVEALTNFLNGITPIVPLRGITISAGDLSPLSTYIAALITGHPRSKY 240
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      174 NSLTRGHSNVRVLVVEALTNFLNGITPIVPLRGITISAGDLSPLSTYIAALISGHPDSKY 233
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      241 HXHXGEXKIMKARBAIALFGLPEVYLGPKEGLGVNGTAVSASMATIALHDAHMLSLLS 300
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      234 HXHXGEXKILYARBAALFNLPEVYLGPKGLGVNGTAVSASMATIALHDAHMLSLLS 293
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      301 QALLTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVANRIRLLGSAFVHHEEVYKVD 360
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVANRIRLLGSAFVHHEEVYKVD 353
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      361 DSGILRODRYPARTSPQWGLPVSDMIHAHVLISLEAGOSTTNDPLDIVENKXTHHGDNF 420
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      354 DSGILRODRYPARTSPQWGLPVSDLIHAHVLITIEAGOSTTNDPLDIVENKXTHHGDNF 413
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      421 QASAVXNTEKTRIALALIGKLNFTQLTETMLNAGNRLPSCLAEDPSLSYHCKGLDIA 480
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      414 QAAAVANTMEKTRILGAIQIKLNFQLTETMLNAGNRLPSCLAEDPSLSYHCKGLDIA 473
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      481 AAAYTSELGHLNPTTYTHQPAEMNGQAVNSIALISARTTAENDVLSLLATLHYCVLQ 540
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      474 AAAYTSELGHLNPTTYTHQPAEMANQANSLALISARTTESNDVLSLLATLHYCVLQ 533
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      541 AAVLRAMEEFKKQFPKXXXXLLXQHFQXXXXXXKXKXKXKLEQTNSTYDL 600
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      534 AIDLRAXEFFFKKQGPALVSLIDQHFSGAMGSLRDELVEKNKTLAKLEQTNSTYDL 593
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      601 EPRMHDASFAXATGVVEKLSXXXXXVSLAAVNAKVAAXEYALSTRVXXKFWAASS 660
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB      594 VPRMHDASFAGTGVVEVLS---TSLSLAAVNAKVAAXEYALSTRQVRETFWGAAST 650
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      661 SSPALXYSPRTVLYSPVEELGVKARQGVFLGKQEVITGSVSRITPAISGRINXY 720
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      651 SSPALXYSPRTQILYAFVVEELGVKARRQGVFLGKQEVITGSVSRITPAISGRINNV 710
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      721 LVKMLA 726
        | | | | |
DB      711 LTKMLA 716
        | | | | |

RESULT 8
US-09-765-873A-35
; Sequence 35, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 35
; LENGTH: 716
; TYPE: PRP
; ORGANISM: mutant from Rhodocorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X = Asp, Asn, Glu, Gln
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X = Thr, Ala, Ser, Pro, Gly
US-09-765-873A-35

Query Match      89.5%; Score 3012; DB 4; Length 716;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

QY      1 MAPSLDSTISXANGXNKGKHAAXAAXXXXXXXXAGXSLTTYXTQLDIYEXXLD 60
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1 MAPSLDSTISHSFANGVSAKQAVNGAS-----TNLAAGSHLPTTYQVDIYEXKMLAA 54
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      61 PXTDXXELDGYSLTLGQVVGAAKRGKXVRYXDSDEIRXKIDKSVFPRQIXNSVYGT 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      55 P-TDSTLELDGYSNLGQVVSAAKRGKRPVRYKDSDEIRSKIDKSVFPRQIXNSVYGT 113
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      121 TGFSGSADTRTEDAISLOKALLEHQLCGVLPFSXDSFXLGRLENSLPLEVVGAMTIRV 180
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      114 TGFSGSADTRTEDAISLOKALLEHQLCGVLPFSXDSFXLGRLENSLPLEVVGAMTIRV 173
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      181 NSLTRGHSNVRVLVVEALTNFLNGITPIVPLRGITISAGDLSPLSTYIAALITGHPRSKY 240
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      174 NSLTRGHSNVRVLVVEALTNFLNGITPIVPLRGITISAGDLSPLSTYIAALISGHPDSKY 233
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      241 HXHXGEXKIMKARBAIALFGLPEVYLGPKEGLGVNGTAVSASMATIALHDAHMLSLLS 300
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      234 HXHXGEXKILYARBAALFNLPEVYLGPKGLGVNGTAVSASMATIALHDAHMLSLLS 293
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      301 QALLTALTYEAMVGHAGSFHPLHDVTRPHPTQIEVANRIRLLGSAFVHHEEVYKVD 360
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      294 QSLTAMTYEAMVGHAGSFHPLHDVTRPHPTQIEVANRIRLLGSAFVHHEEVYKVD 353
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      361 DSGILRODRYPARTSPQWGLPVSDMIHAHVLISLEAGOSTTNDPLDIVENKXTHHGDNF 420
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      354 DSGILRODRYPARTSPQWGLPVSDLIHAHVLITIEAGOSTTNDPLDIVENKXTHHGDNF 413
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      421 QASAVXNTEKTRIALALIGKLNFTQLTETMLNAGNRLPSCLAEDPSLSYHCKGLDIA 480
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
Db 414 QAAAVANTMEKTRGLAIGKLNFTQLTMLNAGNRRGLPSCIAEDPSLSHCKGLDIA 473
Qy 481 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANDVLSLLATHLYCVLQ 540
Db 474 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANDVLSLLATHLYCVLQ 533
Qy 541 AVDLAMEFEFKQEPKXXXXLXQHFQXXXXXXELXXXKXKXLEQTSYDL 600
Db 534 AIDLRAXEPEFKQEPKPAIVSLIDQHFQSGAMTGSNLRDELVEKVKTKAKLEQTSYDL 593
Qy 601 EPRHDAFSXATGTVVEXLSSXXXXVSLAAVNAWKVAAEKALISTEXVXXFMPXAPSS 660
Db 594 VPRHDAFSXATGTVVEXLSS--TSLSLAAVNAWKVAAESALISLTQVRETFWMAAST 650
Qy 661 SSPALXYLSPTRTVLVSFVAREELGVKARRGDVFLGKQEVLTGSNVSRIYEAIKSGRINXV 720
Db 651 SSPALXYLSPTRTQILVAFVAREELGVKARRGDVFLGKQEVLTGSNVSRIYEAIKSGRINXV 710
Qy 721 LVKMLA 726
Db 711 LKXMLA 716

RESULT 9
US-09-765-873A-36
; Sequence 36, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (181)..(181)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (235)..(235)
; OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-36

Query Match 89.5%; Score 3010; DB 4; Length 716;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 608; Conservative 24; Mismatches 84; Indels 10; Gaps 3;
```

```
Db 174 NSLTRGHSAVRLVLEALTNFLNHGITPVLGRGISASGDLSPSLYIAAALIGHPSKV 233
Qy 241 HVXHGEXEKIXASAAIALFGLEPVLAGPKGLG:VNGTAVASAMATLALDHAMLSLS 300
Db 234 HVXHGEXEKILYARAAALFPLEPVLAGPKGLG:VNGTAVASAMATLALDHAMLSLS 293
Qy 301 QALITALVEAMVGHAGSFHPLADYTRPHPTQIEVANNITLLEGSXFVAVHHEEVKXD 360
Db 294 QSLTAMTEAMVGHAGSFHPLADYTRPHPTQIEVAGNIRKLLEGSRFVAVHHEEVKXD 353
Qy 361 DSGILRORVPLRPSQWLGPLVSDMLHAHNVLSLEGGSTTNPPLIDVENKXTHHGNF 420
Db 354 DSGILRORVPLRPSQWLGPLVSDMLHAHNVLTIEGGSTTNPPLIDVENKXTHHGNF 413
Qy 421 QASAVXNTMEKTRGLAIGKLNFTQLTMLNAGNRRGLPSCIAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRGLAIGKLNFTQLTMLNAGNRRGLPSCIAEDPSLSYHCKGLDIA 473
Qy 481 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANDVLSLLATHLYCVLQ 540
Db 474 AAAYTSELGLANPVTHVQPAEMGNQAVNSLALISARPTAEANDVLSLLATHLYCVLQ 533
Qy 541 AVDLAMEFEFKQEPKXXXXLXQHFQXXXXXXELXXXKXKXLEQTSYDL 600
Db 534 AIDLRAXEPEFKQEPKPAIVSLIDQHFQSGAMTGSNLRDELVEKVKTKAKLEQTSYDL 593
Qy 601 EPRHDAFSXATGTVVEXLSSXXXXVSLAAVNAWKVAAEKALISTEXVXXFMPXAPSS 660
Db 594 VPRHDAFSXATGTVVEXLSS--TSLSLAAVNAWKVAAESALISLTQVRETFWMAAST 650
Qy 661 SSPALXYLSPTRTVLVSFVAREELGVKARRGDVFLGKQEVLTGSNVSRIYEAIKSGRINXV 720
Db 651 SSPALXYLSPTRTQILVAFVAREELGVKARRGDVFLGKQEVLTGSNVSRIYEAIKSGRINXV 710
Qy 721 LVKMLA 726
Db 711 LKXMLA 716

RESULT 10
US-09-765-873A-37
; Sequence 37, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 37
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X= Val, Met, Leu, Cys
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-37

Query Match 89.5%; Score 3010; DB 4; Length 716;
```



```

US-09-765-873A-33
; Sequence 33, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 33
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (126)..(126)
; OTHER INFORMATION: X= Gly, Ser, Ala, Thr
; NAME/KEY: UNSURE
; LOCATION: (138)..(138)
; OTHER INFORMATION: X= Leu, Met, Ile, Val, Cys
; NAME/KEY: UNSURE
; LOCATION: (149)..(149)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (181)..(181)
; OTHER INFORMATION: X= Pro, Ala, Ser, Thr, Gly
; NAME/KEY: UNSURE
; LOCATION: (198)..(198)
; OTHER INFORMATION: X= Asp, Asn, Glu, Gln
; NAME/KEY: UNSURE
; LOCATION: (202)..(202)
; OTHER INFORMATION: X= Val, Met, Leu, Cys
; NAME/KEY: UNSURE
; LOCATION: (235)..(235)
; OTHER INFORMATION: X= Ala, Gly, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (502)..(502)
; OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
; NAME/KEY: UNSURE
; LOCATION: (540)..(540)
; OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-33

```

```

Query Match      88.5%; Score 2978; DB 4; Length 716;
Best Local Similarity 82.9%; Pred. No. 0;
Matches 602; Conservative 25; Mismatches 89; Indels 10; Gaps 3;

```

```

QY 1 MAPSLSIATSGXANGXNGXHAAXXASXXXXXXAAGSXLPTTXXTQLDIVEXXLLAD 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MAPSLSIATSGXANGXNGXHAAXXASXXXXXXAAGSXLPTTXXTQLDIVEXXLLAA 54
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 PXTDDXXELDGYSLTLGDVVGAAKRGKRVYVXDSEIRKXIDKSVFRLXQXNSVYGV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 55 P-TDSTLELDGYSLNLDGVVSAARKGRPVYVXDSEIRKIDKSVFRLSQLSWSVYGV 113
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 TGGGSAADRTEDALSLQALLLEHQLCGVLPISXDSFYLGRGLNSLPLEVVRGAMTIRV 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 114 TGGGSAADRTEDALSLQALLLEHQLCGVLPISXDSFYLGRGLNSLPLEVVRGAMTIRV 173
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYAAAITGHPDSKV 240
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 174 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYAAAITGHPDSKV 233
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 241 HXVHEGEXKIMAREALFGLPEVVLGPKSGGLGVNGTAVSASMATLALHDAMLSLLS 300
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 234 HXVHEGEXKIMAREALFGLPEVVLGPKSGGLGVNGTAVSASMATLALHDAMLSLLS 293
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

QY 301 QALTATVEAMVAGSFPHPFADVTSPHPDTEIVARNITRLLEGSGFAVHEEYKVD 360
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 294 QSLTAMTVEAMVAGSFPHPFADVTSPHPDTEIVARNITRLLEGSGFAVHEEYKVD 353
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 361 DEGITRQDRYLRISPMQLGPIVSDMTAAHAYVLSLEAGOSTTNPILDENKXTHGKGF 420
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 354 DEGITRQDRYLRISPMQLGPIVSDMTAAHAYVLSLEAGOSTTNPILDENKXTHGKGF 413
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 421 QASAVXNTMEKTRIALLLIGKLNFTQLTEMLNMGMRGLPSCLAEDPSSISYCKGLDIA 480
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 414 QASAVXNTMEKTRIALLLIGKLNFTQLTEMLNMGMRGLPSCLAEDPSSISYCKGLDIA 473
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 481 AAAYTSELGHLANPVTTHVQPAENQAVNSLLISARPTAEANDVLSLLATHLYCVIQ 540
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 474 AAAYTSELGHLANPVTTHVQPAENQAVNSLLISARPTAEANDVLSLLATHLYCVIQ 533
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 541 AVDLPAVEFEFKQFPKXXLLXQHRGXXXTXXXXXELXKXVXKXKRLKQNTSYDL 600
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 534 AIDLPAVEFEFKQFPKXXLLXQHRGXXXTXXXXXELXKXVXKXKRLKQNTSYDL 593
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 601 EPRWDAFSXATGTVEVXLSXXXXXVSLAANVAMKVAAXEKAISLTXVYXXFXWAPSS 660
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 594 VPRWDAFSXATGTVEVXLSXXXXXVSLAANVAMKVAAXEKAISLTXVYXXFXWAPSS 650
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 661 SSPALXVSPRTVLYSFVREELGVKARQDVFUGKQVTTGSNVRITIEAISKHIXY 720
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 651 SSPALXVSPRTVLYSFVREELGVKARQDVFUGKQVTTGSNVRITIEAISKHIXY 710
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 721 LYKMLA 726
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 711 LYKMLA 716
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 13
US-09-765-873A-32
; Sequence 32, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 686
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-32

```

```

Query Match      88.3%; Score 2971; DB 4; Length 686;
Best Local Similarity 86.2%; Pred. No. 0;
Matches 595; Conservative 23; Mismatches 68; Indels 4; Gaps 2;

```

```

QY 37 AAGSXLPTTXXTQLDIVEXXLLADPXTDDXXELDGYSLTLGDVVGAAKRGKRVYVXDSE 96
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 1 AAGSXLPTTXXTQLDIVEXXLLADPXTDDXXELDGYSLTLGDVVGAAKRGKRVYVXDSE 59
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 97 IRXKIDKSVFRLXQXNSVYGVTTGGGSAADRTEDALSLQALLLEHQLCGVLPISXDS 156
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 60 IRXKIDKSVFRLXQXNSVYGVTTGGGSAADRTEDALSLQALLLEHQLCGVLPISXDS 119
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 157 FYLGRGLNSLPLEVVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGITPIVPLRGIT 216
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 120 FYLGRGLNSLPLEVVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGITPIVPLRGIT 179
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 217 SASGDLSPSYIAAATGHPDSKVYVHEGEXKIMAREALFGLPEVVLGPKSGGLGV 276
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```
Db 180 SASGDLSPSYIAAASGHPDSKVHVHEGKELLYAREAMALFNLEBPVYLGPKEGLV 239
QY 277 NGTAVASAMATLALDPAHMLSLISQALATLVEAMVGHAGSPHPLHDVTPHPTQLEVA 336
Db 240 NGTAVASAMATLALDPAHMLSLISQALATLVEAMVGHAGSPHPLHDVTPHPTQLEVA 299
QY 337 RNIRLLSGSXAFAVHHEEVKVKDEGILLRQDRYPLRTSPQWGLPVSMDTHAAVLSLE 396
Db 300 GNIRKLLEGSRAFAVHHEEVKVKDEGILLRQDRYPLRTSPQWGLPVSMDTHAAVLSLE 359
QY 397 AGOSTDNENLIVENKXTHHGNFQASAVXMTMETRLALIGLANTQTLEMLNAGM 456
Db 360 AGOSTDNENLIVENKXTHHGNFQASAVXMTMETRLALIGLANTQTLEMLNAGM 419
QY 457 RGLPSCIAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTHVQPAEMGNQAVNSLALIS 516
Db 420 RGLPSCIAAEDPSLSYHCKGLDIAAAAYTSELGHLANPVTHVQPAEMGNQAVNSLALIS 479
QY 517 ARTTAEANDVLELLATHLYCVLQAVDLPAHEFEKQFQXKXKXLLXQHFQKXKXKX 576
Db 480 ARTTESNDVLELLATHLYCVLQAVDLPAHEFEKQFQXKXKXLLXQHFQKXKXKX 539
QY 577 XXELKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 636
Db 540 RQELVEKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 596
QY 637 VAXAEKALISLTXVXXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 696
Db 597 VAAAEKALISLTXVXXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 656
QY 697 QEVTTGSNVRIVEAIKSGRLNVLVKMLA 726
Db 657 QEVTTGSNVRIVEAIKSGRLNVLVKMLA 686

RESULT 14
US-09-624-693A-17
; Sequence 17, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koester, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 17
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Rhodotorula mucilaginosa
US-09-624-693A-17

Query Match 86.8%; Score 2922.5; DB 4; Length 713;
Best Local Similarity 82.0%; Pred. No. 0;
Matches 555; Conservative 33; Mismatches 85; Indels 13; Gaps 5;

QY 1 MAPSLDLSATSGXANGXNGHAAAXXASXXXXXXAXAGSKLPTTXXTQLDIVEXXLD 60
Db 1 MAPSVDSIATSGVANSLSNGHAAAXXAN-GSDVNHKKTAGAGSLPTTETQTDIVERTLD 59
QY 61 PXTDXHXELDGLSLTGLGVYGAARGRVYRXDDELTKKIDKSVBELRQOLKNSVYGV 120
Db 60 AGATQIKLIDGYTLTGLGVYGAARGRVYRXDDELTKKIDKSVBELRQOLKNSVYGV 119
QY 121 TGFSGADRTTDAISLQKALIEHQLCVLPTSXDSFYLIGLENSPLPLEVGRGAMTIRV 180
Db 120 TGFSGADRTTDAISLQKALIEHQLCVLPTSMDFLISGLGLENPLPLEVGRGAMTIRV 179
QY 181 NSLIRGHSVAVRLVLEALTNFLNHGITPIVPLRGITISASGDLSPSYIAAATGHPDSKV 240
```

```
Db 180 NSLIRGHSVAVRLVLEALTNFLNHGITPIVPLRGITISASGDLSPSYIAAATGHPDSKV 239
QY 241 HYHGXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 300
Db 240 HV--DG--KIMSQAIALKGLQPVVLGPKEGLVNGTAVASAMATLALDPAHMLSL 295
QY 301 QALTATVEAMVGHAGSPHPLHDVTPHPTQIEVARNIRTLGSGXFAVHHEEVKVKD 360
Db 296 QALTATVEAMVGHAGSPHPLHDVTPHPTQIEVARNIRTLGSGXFAVHHEEVKVKD 355
QY 361 DEGIIRQDRYPLRTSPQWGLPVSMDTHAAVLSLEAGOSTTNDPLIDVENKXTHHGNF 420
Db 356 DEGIIRQDRYPLRTSPQWGLPVSMDTHAAVLSLEAGOSTTNDPLIDVENKXTHHGNF 415
QY 421 QASAVXMTMETRLALIGLANTQTLEMLNAGMNLBSCIAAEDPSLSYHCKGLDIA 480
Db 416 MASSVGNMTMETRLALIGLANTQTLEMLNAGMNLBSCIAAEDPSLSYHCKGLDIA 475
QY 481 AAAYTSELGHLANPVTHVQPAEMGNQAVNSLALISARTTAEANDVLELLATHLYCVLQ 540
Db 476 AAAYTSELGHLANPVTHVQPAEMGNQAVNSLALISARTTAEANDVLELLATHLYCVLQ 535
QY 541 AVDLPAHEFEKQFQXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
Db 536 AVDLPAHEFEKQFQXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 590
QY 601 EPRMWDASXATGTVEXLSXXXXXVSLAAVNAKVAAXAEKALISLTXVXXKXKXKXKXKXKX 660
Db 591 EPRMWDTSVATGAVVELAG---QEVSLASLNAMKVAEAEKALISLTXVXXKXKXKXKXKXKX 647
QY 661 SSPALXYLSPRTVLYSFVREELGVKARGDVFLGQEVTTGSNVRIVEAIKSGRLNVLVKMLA 720
Db 648 SSPALXYLSPRTVLYSFVREELGVKARGDVFLGQEVTTGSNVRIVEAIKSGRLNVLVKMLA 707
QY 721 LVKMLA 726
Db 708 LVKMLA 713

RESULT 15
US-09-624-693A-13
; Sequence 13, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Koester, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 294/9/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 13
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Rhodotorula graminis
US-09-624-693A-13

Query Match 82.3%; Score 2770; DB 4; Length 720;
Best Local Similarity 77.5%; Pred. No. 4; 6e-309;
Matches 563; Conservative 40; Mismatches 117; Indels 6; Gaps 3;

QY 1 MAPSLDLSATSGXANGXNGHAAAXXASXXXXXXAXAGSKLPTTXXTQLDIVEXXLD 60
Db 1 MAPSLDLSATSGXANGXNGHAAAXXASXXXXXXAXAGSKLPTTXXTQLDIVEXXLD 60
QY 61 PXTDXHXELDGLSLTGLGVYGAARGRVYRXDDELTKKIDKSVBELRQOLKNSVYGV 120
Db 61 P-TDVELSLGSLTVRVDVGAARGRVYRXDDELTKKIDKSVBELRQOLKNSVYGV 119
```


QY 121 TGFSGSADTRTEDAISLOKALLEHOLCGVLPISXDSPXILGRLENSLPLEVVRGAMTIRV 180
 Db 120 TGFSGSADTRTEDAISLOKALLEHOLCGVLPISVSSFSVGRGLENTLPLEVVRGAMTIRV 179
 QY 181 NSLTRGHSARLVVLEALTNPLNHGTPIVPRGTISASGDLSPISYIAAITGHPDSKY 240
 Db 180 NSLTRGHSARLVVLEALTNPLNHGTPIVPRGTISASGDLSPISYIAGITGHPDKV 239
 QY 241 HVHGEKEXIMAREALFGLPEVYLGPKEGLVNGTAVASAMATLALHDAHMLSL 300
 Db 240 HVHGEKEXIMAREALFGLPEVYLGPKEGLVNGTAVASAMATLALHDAHMLSL 299
 QY 301 QALTALTVEAMNGHAGSFHPLHDVTRPHPTQIEVARNIPTLBSKFAVHHEEVKXD 360
 Db 300 QALTALTVEAMNGHAGSFHPLHDVTRPHPTQIEVARNIPTLBSKSFVHEHEEVKXD 359
 QY 361 DEGILRDRYPLRTSPQWLGPLVSDMIAHNAVLSLEAGOSTDNPDIIVENKXTHHGNF 420
 Db 360 DEGILRDRYPLRTSPQWLGPLVSDMIAHNAVLSLEAGOSTDNPDIIVENKXTHHGNF 418
 QY 421 QASAVXNTEKTRIALALIGKLNFTQLEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
 Db 419 QASAVSISMETRLALALIGKLNFTQCTELNNAANNRGLPSCLAEDPSLNYHCKGLDIH 478
 QY 481 AAAYTSEIGHLANPVTHVOPAEKNGOAVNSLALISARRTAEANDVLSLLATHLVCVLO 540
 Db 479 IAAVASELGHLANPVTHVOPAEKNGOAVNSLALISARRTAEANDVLSLLASHLYCTLQ 538
 QY 541 AVDLRAMEFEFKQFPXXXXLLOHFGXXXXXXE:XXKVKXLKRLBQJNSYDL 600
 Db 539 AVDLRAMELDFKKQFPPLPTLLQOHLGTGLDVNA---LALBVKKALNKRLBQJTTYDL 594
 QY 601 EPRNHDAFSXATGTVEKLSXXXXXVSLAANANKVAXAEKALSLTRXVXXFWXAPSS 660
 Db 595 EPRNHDAFSYATGTVEKLSXSSPSANVTLTAVNANKVASAEKALSLTREVRNRFWQTPSS 654
 QY 661 SSPALXYLSPRFVLSFVREELGVARRGDVFLGKQEVTTIGSNVSRITYEAIKSGRLNXV 720
 Db 655 QAPAHAYLSPRTRVLXSFVREELGVARRGDVFGVQOETIGSNVSRITYEAIKSGRLNHV 714
 QY 721 LVFMMA 726
 Db 715 LVFMMA 720

Search completed: September 9, 2004, 10:09:41
 Job time : 26.5975 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 10:07:54 ; Search time 93.3859 Seconds
(without alignments)
2493.103 Million cell updates/sec

Title: US-09-939-408a-21

Sequence: 1 MAPSDLSATSTXANGXNGX.....RIYBAKSGRINVLVYKMLA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 32068961 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3223	95.8	726	9	US-09-939-408a-21
2	3021	89.8	716	9	US-09-765-873a-8
3	3021	89.8	716	9	US-09-939-408a-19
4	3021	89.8	716	14	US-10-138-970a-4
5	3021	89.8	716	14	US-10-168-523b-8
6	3021	89.8	716	15	US-10-374-366-2
7	3021	89.8	716	15	US-10-439-478-2
8	3019	89.7	716	9	US-09-765-873a-10
9	3019	89.7	716	14	US-10-138-970a-24
10	3019	89.7	716	14	US-10-168-523b-10
11	3019	89.7	716	15	US-10-374-366-4
12	3016	89.6	716	15	US-10-374-366-24
13	3015	89.6	716	15	US-10-374-366-22
14	3014	89.6	716	9	US-09-765-873a-38
15	3014	89.6	716	14	US-10-138-970a-22

16	3014	89.6	716	14	US-10-168-523b-38	Sequence 38, Appl
17	3014	89.6	716	15	US-10-374-366-16	Sequence 16, Appl
18	3014	89.6	716	15	US-10-439-478-18	Sequence 18, Appl
19	3013	89.5	716	15	US-10-374-366-8	Sequence 8, Appl
20	3012	89.5	716	9	US-09-765-873a-35	Sequence 19, Appl
21	3012	89.5	716	14	US-10-138-970a-19	Sequence 35, Appl
22	3012	89.5	716	14	US-10-168-523b-35	Sequence 35, Appl
23	3012	89.5	716	15	US-10-374-366-10	Sequence 10, Appl
24	3012	89.5	716	15	US-10-439-478-15	Sequence 15, Appl
25	3011	89.5	716	15	US-10-374-366-14	Sequence 14, Appl
26	3010	89.5	716	9	US-09-765-873a-36	Sequence 36, Appl
27	3010	89.5	716	9	US-09-765-873a-37	Sequence 37, Appl
28	3010	89.5	716	14	US-10-138-970a-20	Sequence 20, Appl
29	3010	89.5	716	14	US-10-138-970a-21	Sequence 21, Appl
30	3010	89.5	716	14	US-10-168-523b-36	Sequence 36, Appl
31	3010	89.5	716	14	US-10-168-523b-37	Sequence 37, Appl
32	3010	89.5	716	15	US-10-439-478-16	Sequence 16, Appl
33	3010	89.5	716	15	US-10-439-478-17	Sequence 17, Appl
34	3006	89.3	716	9	US-09-765-873a-34	Sequence 34, Appl
35	3006	89.3	716	14	US-10-138-970a-18	Sequence 18, Appl
36	3006	89.3	716	14	US-10-168-523b-34	Sequence 34, Appl
37	3006	89.3	716	15	US-10-374-366-18	Sequence 18, Appl
38	3006	89.3	716	15	US-10-439-478-14	Sequence 14, Appl
39	3005	89.3	716	15	US-10-374-366-6	Sequence 6, Appl
40	3005	89.3	716	15	US-10-374-366-20	Sequence 20, Appl
41	2997	89.1	716	15	US-10-374-366-12	Sequence 12, Appl
42	2978	88.5	716	9	US-09-765-873a-33	Sequence 33, Appl
43	2978	88.5	716	14	US-10-138-970a-17	Sequence 17, Appl
44	2978	88.5	716	14	US-10-168-523b-33	Sequence 33, Appl
45	2978	88.5	716	15	US-10-439-478-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-939-408a-21
; Sequence 21, Application US/09939408A
; Patent No. US20020102712A1
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; TITLE OF INVENTION: Poly(Alanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; FILE REFERENCE: 29479/500NSCA
; CURRENT APPLICATION NUMBER: US/09/939, 408A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/624, 693
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: PCT/US01/23270
; PRIOR FILING DATE: 2001-07-24
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)..(719)
; OTHER INFORMATION: "Yaa"means any amino acid; "Yaa"means no consensus at that
; OTHER INFORMATION: Description of Artificial Sequence: Consensus of
US-09-939-408a-21

Query Match 95.8%; Score 3223; DB 9; Length 726;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 726; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPSDLSATSTXANGXNGXAAAXXXXXXXXXXXXXXAGSLPTTXXTQDLIVEXXLIAD 60
DB 1 MAPSDLSATSTXANGXNGXAAAXXXXXXXXXXXXXXAGSLPTTXXTQDLIVEXXLIAD 60

QY 61 PXTDDXXELDGYSLTIGDVVGAARKGRVXVDSDEIRKIDKSVFELRXOLXNSVYGV 120
 DB 61 PXTDDXXELDGYSLTIGDVVGAARKGRVXVDSDEIRKIDKSVFELRXOLXNSVYGV 120
 QY 121 TGFSGSADRTEDAIISLOKALLEHOLCGVLPISXSPFXLGRGLENSLEFVVRGAMTIRV 180
 DB 121 TGFSGSADRTEDAIISLOKALLEHOLCGVLPISXSPFXLGRGLENSLEFVVRGAMTIRV 180
 QY 181 NSLTRGHSARVLYVLEALTNFLNHGTPVPRGTTISASGDLSPSTYIAAITHGPPSKV 240
 DB 181 NSLTRGHSARVLYVLEALTNFLNHGTPVPRGTTISASGDLSPSTYIAAITHGPPSKV 240
 QY 241 HYNHEGEXEKIMKAREAIALFGLBEPVVLGPKEGGLVNGTAVASASMATLALDHAMLSLS 300
 DB 241 HYNHEGEXEKIMKAREAIALFGLBEPVVLGPKEGGLVNGTAVASASMATLALDHAMLSLS 300
 QY 301 QALTALTVAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 DB 301 QALTALTVAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 QY 361 DEGLIRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGOSTTDNPLIDVENKXTHHGNF 420
 DB 361 DEGLIRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGOSTTDNPLIDVENKXTHHGNF 420
 QY 421 QASAVXNTMEKTRIALALIGKLNFTQUTEMLNAGMNRGLPSCLAEDPSSLSTHCKGLDIA 480
 DB 421 QASAVXNTMEKTRIALALIGKLNFTQUTEMLNAGMNRGLPSCLAEDPSSLSTHCKGLDIA 480
 QY 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLYCVLQ 540
 DB 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLYCVLQ 540
 QY 541 AVDLRAMEFEFKQFPXKXLLXOHFGKXKTXKXKXELXKXKXKXKLEQTNSTYDL 600
 DB 541 AVDLRAMEFEFKQFPXKXLLXOHFGKXKTXKXKXELXKXKXKXKLEQTNSTYDL 600
 QY 601 EPRWHDASFSAATGTVEKLSKXXXXXVSLAANAKVAXAEKASISLTRYXXKFWXAPSS 660
 DB 601 EPRWHDASFSAATGTVEKLSKXXXXXVSLAANAKVAXAEKASISLTRYXXKFWXAPSS 660
 QY 661 SSPALXYLSPRTVLVSFVREELGVKARGDVFGLGKEVTTIGSNVSRITYEAIKSGRINXV 720
 DB 661 SSPALXYLSPRTVLVSFVREELGVKARGDVFGLGKEVTTIGSNVSRITYEAIKSGRINXV 720
 QY 721 LVKOLA 726
 DB 721 LVKOLA 726
 RESULT 2
 US-09-765-873A-8
 ; Sequence 8, Application US/09765873A
 ; Patent No. US20010053847A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Xiao-Song
 ; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
 ; FILE REFERENCE: BC1009 US CIP
 ; CURRENT APPLICATION NUMBER: US/09/765, 873A
 ; PRIOR FILING DATE: 2001-01-19
 ; PRIOR APPLICATION NUMBER: US 09/627, 216
 ; PRIOR FILING DATE: 2000-07-27
 ; PRIOR APPLICATION NUMBER: US 60/147, 719
 ; PRIOR FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 8
 ; LENGTH: 716
 ; TYPE: PRT
 ; ORGANISM: Rhodocorula glutinis
 US-09-765-873A-8
 Query Match 89.8%; Score 3021; DB 9; Length 716;

Best Local Similarity 84.0%; Pred. No. 7,1e-298;
 Matches 610; Conservaty 25; Mismatches 81; Indels 10; Gaps 3;
 QY 1 NAPSIDSIATSTXANGKXNGKHAAXXAXXXXXXAXXAXXKPTTYXTTOLDIVEXKAD 60
 DB 1 NAPSIDSIATSTXANGKXNGKHAAXXAXXXXXXAXXAXXKPTTYXTTOLDIVEXKAD 60
 QY 55 P-TDSTLBDGYSNLGCVVSAARKGRPVYKSDERISKIDSVFELRQGLMSYVGT 113
 DB 55 P-TDSTLBDGYSNLGCVVSAARKGRPVYKSDERISKIDSVFELRQGLMSYVGT 113
 QY 121 TGFSGSADRTEDAIISLOKALLEHOLCGVLPISXSPFXLGRGLENSLEFVVRGAMTIRV 180
 DB 121 TGFSGSADRTEDAIISLOKALLEHOLCGVLPISXSPFXLGRGLENSLEFVVRGAMTIRV 180
 QY 181 NSLTRGHSARVLYVLEALTNFLNHGTPVPRGTTISASGDLSPSTYIAAITHGPPSKV 240
 DB 181 NSLTRGHSARVLYVLEALTNFLNHGTPVPRGTTISASGDLSPSTYIAAITHGPPSKV 240
 QY 241 HYNHEGEXEKIMKAREAIALFGLBEPVVLGPKEGGLVNGTAVASASMATLALDHAMLSLS 300
 DB 241 HYNHEGEXEKIMKAREAIALFGLBEPVVLGPKEGGLVNGTAVASASMATLALDHAMLSLS 300
 QY 301 QALTALTVAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 DB 301 QALTALTVAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSKFAVHHEEVKVD 360
 QY 361 DEGLIRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGOSTTDNPLIDVENKXTHHGNF 420
 DB 361 DEGLIRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGOSTTDNPLIDVENKXTHHGNF 420
 QY 421 QASAVXNTMEKTRIALALIGKLNFTQUTEMLNAGMNRGLPSCLAEDPSSLSTHCKGLDIA 480
 DB 421 QASAVXNTMEKTRIALALIGKLNFTQUTEMLNAGMNRGLPSCLAEDPSSLSTHCKGLDIA 480
 QY 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLYCVLQ 540
 DB 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARRTAEADVLSLLATLHLYCVLQ 540
 QY 541 AVDLRAMEFEFKQFPXKXLLXOHFGKXKTXKXKXELXKXKXKXKLEQTNSTYDL 600
 DB 541 AVDLRAMEFEFKQFPXKXLLXOHFGKXKTXKXKXELXKXKXKXKLEQTNSTYDL 600
 QY 601 EPRWHDASFSAATGTVEKLSKXXXXXVSLAANAKVAXAEKASISLTRYXXKFWXAPSS 660
 DB 601 EPRWHDASFSAATGTVEKLSKXXXXXVSLAANAKVAXAEKASISLTRYXXKFWXAPSS 660
 QY 661 SSPALXYLSPRTVLVSFVREELGVKARGDVFGLGKEVTTIGSNVSRITYEAIKSGRINXV 720
 DB 661 SSPALXYLSPRTVLVSFVREELGVKARGDVFGLGKEVTTIGSNVSRITYEAIKSGRINXV 720
 QY 721 LVKOLA 726
 DB 721 LVKOLA 726
 RESULT 3
 US-09-939-408a-19
 ; Sequence 19, Application US/09939408A
 ; Patent No. US20020102712A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yoshida, Roberta
 ; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
 ; FILE REFERENCE: 29479/500NSCA
 ; CURRENT APPLICATION NUMBER: US/09/939, 408A
 ; PRIOR FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 09/624, 693
 ; PRIOR FILING DATE: 2000-07-24
 ; PRIOR APPLICATION NUMBER: PCT/US01/23270
 ; PRIOR FILING DATE: 2001-07-24

/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 19
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: Rhodotorula toruloides
US-09-939-408A-19

Query Match 89.8%; Score 3021; DB 9; Length 716;
Best Local Similarity 84.0%; Pred. No. 7.1e-298;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

```
QY 1 MAPSIDSIATXANGXNGXHAAXXAXXXXXXXXAXAGSXLPTTXXTQDIDVEXXLA 60
D 1 MAPSIDSIATXANGXNGXHAAXXAXXXXXXXXAXAGSXLPTTXXTQDIDVEXXLA 54
QY 61 PXTDXXLDDGYSLTLGDVGAARKGRXVYXDSDEIRKIDKSVFRLKQLKNSVYGV 120
D 55 P-TDSTLELDGYSLTLGDVGAARKGRVYXDSDEIRKIDKSVFRLKQLKNSVYGV 113
QY 121 TGFSGSADRTEDAI SLQKALLEHQLCGVLPTXSDFYLGRLGENSELPLEVRGAMTIRV 180
D 114 TGFSGSADRTEDAI SLQKALLEHQLCGVLPTXSDFYLGRLGENSELPLEVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGTTISASGDLSPLSYIAAATGHPDSKV 240
D 174 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGTTISASGDLSPLSYIAAATGHPDSKV 233
QY 241 HXHEGKXIKKXARBAIALFGLSEPVVLGPRKGLGVNGTAVSASMATLAHDAMLSLS 300
D 234 HXHEGKXIKKXARBAIALFGLSEPVVLGPRKGLGVNGTAVSASMATLAHDAMLSLS 293
QY 301 QALTALTYEAVVAGSGFHPFLHDVTRPHTQIEVARNIRTLLEGSAFYAVHHEEVYK 360
D 294 QSLTAMTYEAVVAGSGFHPFLHDVTRPHTQIEVARNIRTLLEGSAFYAVHHEEVYK 353
QY 361 DEGLIRORRYPLRTSPQWLGPIVSDMTHAAVLSLEAGQSTTNPPLIDVENKXTHGNGF 420
D 354 DEGLIRORRYPLRTSPQWLGPIVSDMTHAAVLSLEAGQSTTNPPLIDVENKXTHGNGF 413
QY 421 QASAVXNTMEKTRIALALIGKLNFTQLEMLNAGMNGRLPSCLAADPSSYCHCKGLDIA 480
D 414 QAAVANTMEKTRIALALIGKLNFTQLEMLNAGMNGRLPSCLAADPSSYCHCKGLDIA 473
QY 481 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAEANDVLSLLATHLVCVLQ 540
D 474 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAEANDVLSLLATHLVCVLQ 533
QY 541 AVDLRAMEFEFKQFPXXXXLXQHGXXXXTXXXXXXELKXKXKXKXRLKRLKQNTSYDL 600
D 534 AIDLRALIEFEFKQFPXXXXLXQHGXXXXTXXXXXXELKXKXKXKXRLKRLKQNTSYDL 593
QY 601 EPRWDAFSXATGTVEXLSXXXXXVSLAAVNAWKVAAAEKALISLTRVYXXFXMAPSS 660
D 594 VPRWDAFSXATGTVEXLSXXXXXVSLAAVNAWKVAAAEKALISLTRVYXXFXMAPSS 650
QY 661 SSPALXYISPTRTVLVYFVREELGVKARQDVFLGKQEVTTIGSNVSIYBAISGRINXY 720
D 651 SSPALXYISPTRTVLVYFVREELGVKARQDVFLGKQEVTTIGSNVSIYBAISGRINXY 710
QY 721 LVKMLA 726
D 711 LKMLA 716
```

RESULT 4
US-10-138-970A-4
/ Sequence 4, Application US/10138970A
/ Publication No. US20030079255A1
/ GENERAL INFORMATION:
/ APPLICANT: Sartisiani, Sina
/ APPLICANT: Tang, Xiao-Song
/ APPLICANT: Qi, Wei Wei

/ TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
/ FILE REFERENCE: CIL777
/ CURRENT APPLICATION NUMBER: US/10/138, 970A
/ CURRENT FILING DATE: 2002-07-23
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 4
/ LENGTH: 716
/ TYPE: PRT
/ ORGANISM: C. violaceum
US-10-138-970A-4

Query Match 89.8%; Score 3021; DB 14; Length 716;
Best Local Similarity 84.0%; Pred. No. 7.1e-298;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

```
QY 1 MAPSIDSIATXANGXNGXHAAXXAXXXXXXXXAXAGSXLPTTXXTQDIDVEXXLA 60
D 1 MAPSIDSIATXANGXNGXHAAXXAXXXXXXXXAXAGSXLPTTXXTQDIDVEXXLA 54
QY 61 PXTDXXLDDGYSLTLGDVGAARKGRXVYXDSDEIRKIDKSVFRLKQLKNSVYGV 120
D 55 P-TDSTLELDGYSLTLGDVGAARKGRVYXDSDEIRKIDKSVFRLKQLKNSVYGV 113
QY 121 TGFSGSADRTEDAI SLQKALLEHQLCGVLPTXSDFYLGRLGENSELPLEVRGAMTIRV 180
D 114 TGFSGSADRTEDAI SLQKALLEHQLCGVLPTXSDFYLGRLGENSELPLEVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGTTISASGDLSPLSYIAAATGHPDSKV 240
D 174 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGTTISASGDLSPLSYIAAATGHPDSKV 233
QY 241 HXHEGKXIKKXARBAIALFGLSEPVVLGPRKGLGVNGTAVSASMATLAHDAMLSLS 300
D 234 HXHEGKXIKKXARBAIALFGLSEPVVLGPRKGLGVNGTAVSASMATLAHDAMLSLS 293
QY 301 QALTALTYEAVVAGSGFHPFLHDVTRPHTQIEVARNIRTLLEGSAFYAVHHEEVYK 360
D 294 QSLTAMTYEAVVAGSGFHPFLHDVTRPHTQIEVARNIRTLLEGSAFYAVHHEEVYK 353
QY 361 DEGLIRORRYPLRTSPQWLGPIVSDMTHAAVLSLEAGQSTTNPPLIDVENKXTHGNGF 420
D 354 DEGLIRORRYPLRTSPQWLGPIVSDMTHAAVLSLEAGQSTTNPPLIDVENKXTHGNGF 413
QY 421 QASAVXNTMEKTRIALALIGKLNFTQLEMLNAGMNGRLPSCLAADPSSYCHCKGLDIA 480
D 414 QAAVANTMEKTRIALALIGKLNFTQLEMLNAGMNGRLPSCLAADPSSYCHCKGLDIA 473
QY 481 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAEANDVLSLLATHLVCVLQ 540
D 474 AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAEANDVLSLLATHLVCVLQ 533
QY 541 AVDLRAMEFEFKQFPXXXXLXQHGXXXXTXXXXXXELKXKXKXKXRLKRLKQNTSYDL 600
D 534 AIDLRALIEFEFKQFPXXXXLXQHGXXXXTXXXXXXELKXKXKXKXRLKRLKQNTSYDL 593
QY 601 EPRWDAFSXATGTVEXLSXXXXXVSLAAVNAWKVAAAEKALISLTRVYXXFXMAPSS 660
D 594 VPRWDAFSXATGTVEXLSXXXXXVSLAAVNAWKVAAAEKALISLTRVYXXFXMAPSS 650
QY 661 SSPALXYISPTRTVLVYFVREELGVKARQDVFLGKQEVTTIGSNVSIYBAISGRINXY 720
D 651 SSPALXYISPTRTVLVYFVREELGVKARQDVFLGKQEVTTIGSNVSIYBAISGRINXY 710
QY 721 LVKMLA 726
D 711 LKMLA 716
```

RESULT 5
US-10-188-523B-8
/ Sequence 8, Application US/10188523B

Publication No. US20030170834A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
FILE REFERENCE: BC1009 US DIVICP
CURRENT APPLICATION NUMBER: US/10/188,523B
CURRENT FILING DATE: 2002-10-17
PRIOR APPLICATION NUMBER: US 09/627,216
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: US 60/147,719
PRIOR FILING DATE: 1999-08-06
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 716
TYPE: PRT
ORGANISM: Rhodotorula glutinis
US-10-188-523B-8

Query Match 89.8%; Score 3021; DB 14; Length 716;
Best Local Similarity 84.0%; Pred. No. 7.1e-298;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSLDSTATXANGXNGXHAAXXASXXXXXXXAGSXLPTTXXTQLDIVEXXLAD 60
DB 1 MAPSLDSTATXANGXNGXHAAXXASXXXXXXXAGSXLPTTXXTQLDIVEXXLAD 54
QY 61 PXTDDXELDGYSLITGDVGAARKGRVXVDSDEIRKIDKSVFELRXQJXNSYGVY 120
DB 55 P-TDSTLELDGYSLNGDVGAARKGRPVXKSDERSKIDKSVFELRSQJXNSYGVY 113
QY 121 TGFSGSADRTEDIASLOKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 180
DB 114 TGFSGSADRTEDIASLOKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNGITPIVPLRGTTISASGDLSPSTYIAAATGHPDSKY 240
DB 174 NSLTRGSAVRLVLEALTNFLNGITPIVPLRGTTISASGDLSPSTYIAAATGHPDSKY 233
QY 241 HVHGEKEXIKYAREALFGLPEVVLGPKGGLVNGTAVSASMATLALHDAHMLSLLS 300
DB 234 HVHGEKEXIKYAREALFGLPEVVLGPKGGLVNGTAVSASMATLALHDAHMLSLLS 293
QY 301 QALTALTVBAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSPFAVHHEEVKVD 360
DB 294 QSLTAMTVBAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSPFAVHHEEVKVD 353
QY 361 DEGIILQDRYPLRTSPQWLGPIVSDMIHAHVLISLAGOSTTNDPLIDVENKXTHHGNF 420
DB 354 DEGIILQDRYPLRTSPQWLGPIVSDMIHAHVLISLAGOSTTNDPLIDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRGLALIGLNTFTQITEMNAGNRLPSCIAAEDPSLSYHCKGLDIA 480
DB 414 QASAVXNTEKTRGLALIGLNTFTQITEMNAGNRLPSCIAAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGLANPVTTHVQPAEMNGQAVNSIALISARRTPABNDVLSLLATLHYCVLQ 540
DB 474 AAAYTSELGLANPVTTHVQPAEMNGQAVNSIALISARRTPABNDVLSLLATLHYCVLQ 533
QY 541 AVDLRAMEFEFKQFPKXXXXLXQHFQXXXXXXELXXKXKXKXLEQNTSYDL 600
DB 534 AIDLRAIIEFEFKQFPKPAIVSLIDQHFSGAMTGSNLEDELVEKYNKTLAKLEQNTSYDL 593
QY 601 EPRMHDASFAXTGTVEVLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXKXKXAPSS 660
DB 594 VPRMHDASFAXTGTVEVLS--TSLSLAAVNAKVAAXEKAISLTRQVRETFWMSAST 650
QY 661 SSPALXLYSPRTVLYSFVREBELGVKARBDVFLGKQEVTTIGSNVSHIYEAIKSGRINXV 720
DB 651 SSPALXLYSPRTVLYSFVREBELGVKARBDVFLGKQEVTTIGSNVSHIYEAIKSGRINXV 710
QY 721 LVKMLA 726
|:|:|:|

DB 721 LVKMLA 716

RESULT 6
US-10-374-366-2
Sequence 2, Application US/10374366
Publication No. US20040014085A1
GENERAL INFORMATION:
APPLICANT: Tang, Xiao-Song
APPLICANT: Milano, Joseph
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
FILE REFERENCE: CIL794 US NA
CURRENT APPLICATION NUMBER: US/10/374,366
CURRENT FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: 60/360,279
PRIOR FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 203
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2
LENGTH: 716
TYPE: PRT
ORGANISM: Rhodospiridium glutinis
US-10-374-366-2

Query Match 89.8%; Score 3021; DB 15; Length 716;
Best Local Similarity 84.0%; Pred. No. 7.1e-298;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSLDSTATXANGXNGXHAAXXASXXXXXXXAGSXLPTTXXTQLDIVEXXLAD 60
DB 1 MAPSLDSTATXANGXNGXHAAXXASXXXXXXXAGSXLPTTXXTQLDIVEXXLAD 54
QY 61 PXTDDXELDGYSLITGDVGAARKGRVXVDSDEIRKIDKSVFELRXQJXNSYGVY 120
DB 55 P-TDSTLELDGYSLNGDVGAARKGRPVXKSDERSKIDKSVFELRSQJXNSYGVY 113
QY 121 TGFSGSADRTEDIASLOKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 180
DB 114 TGFSGSADRTEDIASLOKALLEHQLCGVLPSTXDSFXLGRGLNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNGITPIVPLRGTTISASGDLSPSTYIAAATGHPDSKY 240
DB 174 NSLTRGSAVRLVLEALTNFLNGITPIVPLRGTTISASGDLSPSTYIAAATGHPDSKY 233
QY 241 HVHGEKEXIKYAREALFGLPEVVLGPKGGLVNGTAVSASMATLALHDAHMLSLLS 300
DB 234 HVHGEKEXIKYAREALFGLPEVVLGPKGGLVNGTAVSASMATLALHDAHMLSLLS 293
QY 301 QALTALTVBAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSPFAVHHEEVKVD 360
DB 294 QSLTAMTVBAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGSPFAVHHEEVKVD 353
QY 361 DEGIILQDRYPLRTSPQWLGPIVSDMIHAHVLISLAGOSTTNDPLIDVENKXTHHGNF 420
DB 354 DEGIILQDRYPLRTSPQWLGPIVSDMIHAHVLISLAGOSTTNDPLIDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRGLALIGLNTFTQITEMNAGNRLPSCIAAEDPSLSYHCKGLDIA 480
DB 414 QASAVXNTEKTRGLALIGLNTFTQITEMNAGNRLPSCIAAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGLANPVTTHVQPAEMNGQAVNSIALISARRTPABNDVLSLLATLHYCVLQ 540
DB 474 AAAYTSELGLANPVTTHVQPAEMNGQAVNSIALISARRTPABNDVLSLLATLHYCVLQ 533
QY 541 AVDLRAMEFEFKQFPKXXXXLXQHFQXXXXXXELXXKXKXKXLEQNTSYDL 600
DB 534 AIDLRAIIEFEFKQFPKPAIVSLIDQHFSGAMTGSNLEDELVEKYNKTLAKLEQNTSYDL 593
QY 601 EPRMHDASFAXTGTVEVLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXKXKXAPSS 660
DB 594 VPRMHDASFAXTGTVEVLS--TSLSLAAVNAKVAAXEKAISLTRQVRETFWMSAST 650
QY 661 SSPALXLYSPRTVLYSFVREBELGVKARBDVFLGKQEVTTIGSNVSHIYEAIKSGRINXV 720
DB 651 SSPALXLYSPRTVLYSFVREBELGVKARBDVFLGKQEVTTIGSNVSHIYEAIKSGRINXV 710
QY 721 LVKMLA 726
|:|:|:|

```
Db 651 SSPALSYSPRTQILYAFVREELGVKARGDVFLGKQEVITIGSNVSKTYEALISGRINNV 710
QY 721 LYKMLA 726
Db 711 LKXMLA 716

RESULT 7
US-10-439-478-2
; Sequence 2, Application US/10439478
; Publication No. US20040018600A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Co.
; APPLICANT: Ben-Basat, Arle
; APPLICANT: Qi, Wei Wei
; APPLICANT: Sartasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Vanelli, Todd
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxyetyrene
; FILE REFERENCE: C11912
; CURRENT APPLICATION NUMBER: US/10/439,478
; PRIOR FILING DATE: 2003-05-16
; PRIOR APPLICATION NUMBER: 60/383450
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-10-439-478-2

Query Match      89.8%; Score 3021; DB 15; Length 716;
Best Local Similarity 84.0%; Pred. No. 7,1e-298;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSIDSISATXANGXNGXHAAXXAXXXXXXAXAGXSLPTTXXTQLDIVEXXLLAD 60
Db 1 MAPSIDSISHSFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQVTDIVEXKMLA 54
QY 61 PXTDXKXELDDGSLTLGDVYGAARKGRVYXDSDEIRKXIKDKSVEFLRQOLXNSVYGT 120
Db 55 P-TDSTLELDGSLNLGDVVSAAKRGPRVYKDSDEIRSKIKDKSVEFLRQOLXNSVYGT 113
QY 121 TGFSGSADRTREDAISLQKALLEHQLCGVLPTSXDSFXLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADRTREDAISLQKALLEHQLCGVLPSDFSFLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSLSTIAAATGHPDSKY 240
Db 174 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSLSTIAAATGHPDSKY 233
QY 241 HVYHGXEXIKMARERAILFGLPEVVLGPKGELGVNGTAVSASMATLALHDAHMLSLLS 300
Db 234 HVYHGXEXIKLYAREKMALEFLEPVVLGPKGELGVNGTAVSASMATLALHDAHMLSLLS 293
QY 301 QALTALTYEAVVGAAGSFHPLADVTRPHPTQIEVARNIRTLLEGXFAVHHEEVKVD 360
Db 294 QSLTAMTYEAVVGAAGSFHPLADVTRPHPTQIEVARNIRTLLEGXFAVHHEEVKVD 353
QY 361 DSGILRQDYPRLRTPSQWLGPIVSDMTHAAVLSLEAGOSTTNDPLLDVENKXTHHGNF 420
Db 354 DSGILRQDYPRLRTPSQWLGPIVSDMTHAAVLSLEAGOSTTNDPLLDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKINFOTLEMLNAGNRRGLPSCLAADPSLSYHCKGLDIA 480
Db 414 QASAVXNTEKTRIRLGAQIGKINFOTLEMLNAGNRRGLPSCLAADPSLSYHCKGLDIA 473
QY 481 AAAYTSELCHLANPYTTTHVQPAEMGNQAVNSIALISARITAEANDVLSLLIATILYCVLQ 540
Db 474 AAAYTSELCHLANPYTTTHVQPAEMGNQAVNSIALISARITAEANDVLSLLIATILYCVLQ 533
```

```
QY 541 AYDLRAMEFEFKQFPKXXXXLLKQHFQXXXTXXXXXXELXAKYKXKXLRLEQTNVDL 600
Db 534 AIDLRAIEFEFKQFQPAIVSLIDQHFQSAVNTGSNLDDELVEKXKXKLARLEQTNVDL 593
QY 601 EPRMHDAFSXATGTVEVXSSXXXXXVSLAAVNAKMYAAEKAISLTRYXRXFWXAPSS 660
Db 594 VPRMHDAFSXATGTVEVXSS---TSLSLAAVNAKMYAAEKAISLTRYXRFETWMSAAS 650
QY 661 SSPALXYSIPRTVLYSVREELGVKARGDVFLGKQEVITIGSNVSKTYEALISGRINNV 720
Db 651 SSPALSYSPRTQILYAFVREELGVKARGDVFLGKQEVITIGSNVSKTYEALISGRINNV 710
QY 721 LYKMLA 726
Db 711 LKXMLA 716

RESULT 8
US-09-765-873A-10
; Sequence 10, Application US/09765873A
; Patent No. US20010053847A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765, 873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 716
; TYPE: PRT
; ORGANISM: mutant from Rhodotorula glutinis
US-09-765-873A-10

Query Match      89.7%; Score 3019; DB 9; Length 716;
Best Local Similarity 84.0%; Pred. No. 1,1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

QY 1 MAPSIDSISATXANGXNGXHAAXXAXXXXXXAXAGXSLPTTXXTQLDIVEXXLLAD 60
Db 1 MAPSIDSISHSFANGVSAKQAVNGAS-----TNLAVAGSHLPTTQVTDIVEXKMLA 54
QY 61 PXTDXKXELDDGSLTLGDVYGAARKGRVYXDSDEIRKXIKDKSVEFLRQOLXNSVYGT 120
Db 55 P-TDSTLELDGSLNLGDVVSAAKRGPRVYKDSDEIRSKIKDKSVEFLRQOLXNSVYGT 113
QY 121 TGFSGSADRTREDAISLQKALLEHQLCGVLPTSXDSFXLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADRTREDAISLQKALLEHQLCGVLPSDFSFLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSLSTIAAATGHPDSKY 240
Db 174 NSLTRGSAVRLVLEBALTNFLNHGITPIVPLRGITISASGDLSPSLSTIAAATGHPDSKY 233
QY 241 HVYHGXEXIKMARERAILFGLPEVVLGPKGELGVNGTAVSASMATLALHDAHMLSLLS 300
Db 234 HVYHGXEXIKLYAREKMALEFLEPVVLGPKGELGVNGTAVSASMATLALHDAHMLSLLS 293
QY 301 QALTALTYEAVVGAAGSFHPLADVTRPHPTQIEVARNIRTLLEGXFAVHHEEVKVD 360
Db 294 QSLTAMTYEAVVGAAGSFHPLADVTRPHPTQIEVARNIRTLLEGXFAVHHEEVKVD 353
QY 361 DSGILRQDYPRLRTPSQWLGPIVSDMTHAAVLSLEAGOSTTNDPLLDVENKXTHHGNF 420
Db 354 DSGILRQDYPRLRTPSQWLGPIVSDMTHAAVLSLEAGOSTTNDPLLDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKINFOTLEMLNAGNRRGLPSCLAADPSLSYHCKGLDIA 480
Db 414 QASAVXNTEKTRIRLGAQIGKINFOTLEMLNAGNRRGLPSCLAADPSLSYHCKGLDIA 473
```

Db 414 QAAVAANTMEKTRGLAIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 473
Qy 481 AAATSEGLHANLPVTHVQPAEMGNQAVNSLALISARTAEANDVLSLLATHLCYCLO 540
Db 474 AAATSEGLHANLPVTHVQPAEMANOAVNSLALISARTTESNDVLSLLATHLCYCLO 533
Qy 541 AVDLRAMEFEFKQFPKXXXLLKHQFGXXXTXXXXXXELXXXKXKXKLEQTNXYDL 600
Db 534 AIDLRATEFEFKQFPKPAIVSLIDQHFSGAMTGSNLDELVEKNKTLAKRLQTNXYDL 593
Qy 601 EPRWHDAPSKATGTVEVEXLSXXXXXVSLAANAKVAAEKALSLTRVXXFXWAPSS 660
Db 594 VPRWHDAPSKATGTVEVEXLS--TSLSLAAVNAWKVAAESALSLTRQVRETFWMAAST 650
Qy 661 SSPALXYLSPTRTVLYSFVREELGVKARRGVFLGKQEVITGSNVSRIYEAIRSGRINXY 720
Db 651 SSPALXYLSPTRTQILYAFVREELGVKARRGVFLGKQEVITGSNVSRIYEAIRSGRINNV 710
Qy 721 LVKMLA 726
Db 711 LKMLA 716

RESULT 9
US-10-138-970A-24
; Sequence 24, Application US/10138970A
; Publication No. US20030079255A1
; GENERAL INFORMATION:
; APPLICANT: Sartasiani, Sima
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-
; FILE REFERENCE: hydroxycinnamic Acid
; FILE REFERENCE: C11777
; CURRENT APPLICATION NUMBER: US/10/138,970A
; CURRENT FILING DATE: 2002-07-23
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 716
; TYPE: PRT
; ORGANISM: artificial sequence
; OTHER INFORMATION: mutant from rhodotorula glutinis
US-10-138-970A-24

Query Match 89.7%; Score 3019; DB 14; Length 716;
Best Local Similarity 84.0%; Pred. No. 1.1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

Qy 1 MAPSLDSIATSKXNGXKXGHAAXXASXXXXXXXAAGSKLPTTXXTQLDIVEXXKLD 60
Db 1 MAPSLDSISHSFANGVASAKQAVNGAS-----TNLAAVAGSHLPTTQVTVQDIVEKXLLA 54
Qy 61 PYTDDXXELDGYSLTLDGVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLANSYYGYT 120
Db 55 P-TDSTLELDGYSLNLDGVVSAARKGRPVXKDSDEIRSKIDKSVEFLRSQLSMSYYGYT 113
Qy 121 TGFSGSADTRTEDAISLQKALLEHQLCGVLPFSXDSFXLGRGLENSLPLEVVGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCGVLPFSFSLRGLGLENSLPLEVVGAMTIRV 173
Qy 181 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISAGDLSPLSYIAAITSQHPDSKY 240
Db 174 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISAGDLSPLSYIAAITSQHPDSKY 233
Qy 241 HVXHEGEXIKAKAREALFGLPEVVLGPKEGGLGVNGTAVASAMATLALHDAHMLSLLS 300
Db 234 HVXHEGEXIKLVAREAMALFNLPEVVLGPKEGGLGVNGTAVASAMATLALHDAHMLSLLS 293
Qy 301 QALTALTAVAMGCHASPHRPLADVTGRPHPTQIEVARNITRLLESXPAVHHEEYVKD 360
Db 294 QSLTAMTAVAMGCHASPHRPLADVTGRPHPTQIEVAGNIRKLLBESRPAVHHEEYVKD 353

Qy 361 DEGILRORRYPLRTSPQWLGPLVSDMTAAHAVLSLEAGOSTTNDPLIDVENKXTHHGNF 420
Db 354 DEGILRORRYPLRTSPQWLGPLVSDMTAAHAVLTIAGOSTTNDPLIDVENKXTHHGNF 413
Qy 421 QASAVANTMEKTRGLAIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAVAANTMEKTRGLAIGKLNFTQLTMLNAGMNRGLPSCLAEDPSLSYHCKGLDIA 473
Qy 481 AAATSEGLHANLPVTHVQPAEMGNQAVNSLALISARTAEANDVLSLLATHLCYCLO 540
Db 474 AAATSEGLHANLPVTHVQPAEMANOAVNSLALISARTTESNDVLSLLATHLCYCLO 533
Qy 541 AVDLRAMEFEFKQFPKXXXLLKHQFGXXXTXXXXXXELXXXKXKXKLEQTNXYDL 600
Db 534 AIDLRATEFEFKQFPKPAIVSLIDQHFSGAMTGSNLDELVEKNKTLAKRLQTNXYDL 593
Qy 601 EPRWHDAPSKATGTVEVEXLSXXXXXVSLAANAKVAAEKALSLTRVXXFXWAPSS 660
Db 594 VPRWHDAPSKATGTVEVEXLS--TSLSLAAVNAWKVAAESALSLTRQVRETFWMAAST 650
Qy 661 SSPALXYLSPTRTVLYSFVREELGVKARRGVFLGKQEVITGSNVSRIYEAIRSGRINXY 720
Db 651 SSPALXYLSPTRTQILYAFVREELGVKARRGVFLGKQEVITGSNVSRIYEAIRSGRINNV 710
Qy 721 LVKMLA 726
Db 711 LKMLA 716

RESULT 10
US-10-188-523B-10
; Sequence 10, Application US/10188523B
; Publication No. US20030170834A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US DIVCIP
; CURRENT APPLICATION NUMBER: US/10/188,523B
; CURRENT FILING DATE: 2002-10-17
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis mutant
US-10-188-523B-10

Query Match 89.7%; Score 3019; DB 14; Length 716;
Best Local Similarity 84.0%; Pred. No. 1.1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

Qy 1 MAPSLDSIATSKXNGXKXGHAAXXASXXXXXXXAAGSKLPTTXXTQLDIVEXXKLD 60
Db 1 MAPSLDSISHSFANGVASAKQAVNGAS-----TNLAAVAGSHLPTTQVTVQDIVEKXLLA 54
Qy 61 PYTDDXXELDGYSLTLDGVVGAARKGRXVRVXDSDEIRXKIDKSVEFLRXQLANSYYGYT 120
Db 55 P-TDSTLELDGYSLNLDGVVSAARKGRPVXKDSDEIRSKIDKSVEFLRSQLSMSYYGYT 113
Qy 121 TGFSGSADTRTEDAISLQKALLEHQLCGVLPFSXDSFXLGRGLENSLPLEVVGAMTIRV 180
Db 114 TGFSGSADTRTEDAISLQKALLEHQLCGVLPFSFSLRGLGLENSLPLEVVGAMTIRV 173
Qy 181 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISAGDLSPLSYIAAITSQHPDSKY 240
Db 174 NSLTRGHSARLVVLEALTNFLNNGITPIVPLRGITISAGDLSPLSYIAAITSQHPDSKY 233
Qy 241 HVXHEGEXIKAKAREALFGLPEVVLGPKEGGLGVNGTAVASAMATLALHDAHMLSLLS 300


```
Db      223  HVHGEKELIYAREAMALFNLEPVVLGPKEGGLVNGTAVASAMATLADHADMSTLS 293
QY      301  QALTALTEVAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGXSFVAHHEEVVKD 360
Db      294  QSLTMTVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLGSGRFAVHHEEVVKD 353
QY      361  DEGLIRQDRYPLRTSPQWLGPLVSDMIHAHVLSEAGOSTTNDPLIDVENKXTHHGNF 420
Db      354  DEGLIRQDRYPLRTSPQWLGPLVSDMIHAHVLTEAGOSTTNDPLIDVENKXTHHGNF 413
QY      421  QASAVXNTEKTRTLALIGKLNFTQLTLEMNAGNRLPSCLAEDPSLSYHCKGLDIA 480
Db      414  QAAAVANTEMKTRGLAGLIGKLNFTQLTLEMNAGNRLPSCLAEDPSLSYHCKGLDIA 473
QY      481  AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAEANDVLSLLATHLVCVLQ 540
Db      474  AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAESNDVLSLLATHLVCVLQ 533
QY      541  AVDLRAMEFEFKQEPKXKXLLXQHFQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
Db      534  AIDLRAEFERKQEPKXKXLLXQHFQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 593
QY      601  ERMWDASFAGTGVVEVLSKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 660
Db      594  VPRWDASFAGTGVVEVLSKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 650
QY      661  SSPALXYISPRTRVLYSVREELGVKARGDVFGLKQEVITGSVNSIYEAISGRINXY 720
Db      651  SSPALXYISPRTRVLYSVREELGVKARGDVFGLKQEVITGSVNSIYEAISGRINXY 710
QY      721  LVKMLA 726
Db      711  LVKMLA 716

RESULT 11
US-10-374-366-4
; Sequence 4, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: C14794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-4

Query Match      89.7%; Score 3019; DB 15; Length 716;
Best Local Similarity 84.0%; Pred. No. 1,1e-297;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;
```

```
QY      1  MAPSLDISATSYANGXNGXAAAXXASXXXXXXAXAGSLPTTXXTQLDIVEXXLIAD 60
Db      1  MAPSLDISISHFANGVAGAKQAVNGAS-----TNLAVAGSHLPTTQVTVQVDIVEXKMLAA 54
QY      61  PXTDXAXEIDGSLTLDGVVGAARKGRVRYXDSDEIRKXIKDKSVEFLRKQLXNSVYGV 120
Db      55  P-TDSTLELDGSLNLDGVVGAARKGRVRYXDSDEIRKXIKDKSVEFLRKQLXNSVYGV 113
QY      121  TGFSGSADRTEDATSLQKALLHQLCGVLPFSXDSFXLGLGLENSJPLEVVRGAMTIRV 180
Db      114  TGFSGSADRTEDATSLQKALLHQLCGVLPFSXDSFXLGLGLENSJPLEVVRGAMTIRV 173
```

```
QY      181  NSLTRGSAVRLVLEALTNFLEHGTPIVPLRGTTIASGDLSPSLSTIAAATGHPDSKY 240
Db      174  NSLTRGSAVRLVLEALTNFLEHGTPIVPLRGTTIASGDLSPSLSTIAAATGHPDSKY 223
QY      241  HVHGEKELIYAREAMALFNLEPVVLGPKEGGLVNGTAVASAMATLADHADMSTLS 300
Db      224  HVHGEKELIYAREAMALFNLEPVVLGPKEGGLVNGTAVASAMATLADHADMSTLS 293
QY      301  QALTALTEVAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLLEGXSFVAHHEEVVKD 360
Db      294  QSLTMTVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLGSGRFAVHHEEVVKD 353
QY      361  DEGLIRQDRYPLRTSPQWLGPLVSDMIHAHVLSEAGOSTTNDPLIDVENKXTHHGNF 420
Db      354  DEGLIRQDRYPLRTSPQWLGPLVSDMIHAHVLTEAGOSTTNDPLIDVENKXTHHGNF 413
QY      421  QASAVXNTEKTRTLALIGKLNFTQLTLEMNAGNRLPSCLAEDPSLSYHCKGLDIA 480
Db      414  QAAAVANTEMKTRGLAGLIGKLNFTQLTLEMNAGNRLPSCLAEDPSLSYHCKGLDIA 473
QY      481  AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAEANDVLSLLATHLVCVLQ 540
Db      474  AAAYTSELGHLANPYTHVQPAEMNGQAVNSLALISARTAESNDVLSLLATHLVCVLQ 533
QY      541  AVDLRAMEFEFKQEPKXKXLLXQHFQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
Db      534  AIDLRAEFERKQEPKXKXLLXQHFQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 593
QY      601  ERMWDASFAGTGVVEVLSKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 660
Db      594  VPRWDASFAGTGVVEVLSKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 650
QY      661  SSPALXYISPRTRVLYSVREELGVKARGDVFGLKQEVITGSVNSIYEAISGRINXY 720
Db      651  SSPALXYISPRTRVLYSVREELGVKARGDVFGLKQEVITGSVNSIYEAISGRINXY 710
QY      721  LVKMLA 726
Db      711  LVKMLA 716

RESULT 12
US-10-374-366-24
; Sequence 24, Application US/10374366
; Publication No. US20040014085A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Milano, Joseph
; TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS
; FILE REFERENCE: C14794 US NA
; CURRENT APPLICATION NUMBER: US/10/374,366
; CURRENT FILING DATE: 2003-02-26
; PRIOR APPLICATION NUMBER: 60/360,279
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 24
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodospiridium glutinis
US-10-374-366-24

Query Match      89.6%; Score 3016; DB 15; Length 716;
Best Local Similarity 83.9%; Pred. No. 2,3e-297;
Matches 609; Conservative 25; Mismatches 82; Indels 10; Gaps 3;
```

```
QY      1  MAPSLDISATSYANGXNGXAAAXXASXXXXXXAXAGSLPTTXXTQLDIVEXXLIAD 60
Db      1  MAPSLDISISHFANGVAGAKQAVNGAS-----TNLAVAGSHLPTTQVTVQVDIVEXKMLAA 54
QY      61  PXTDXAXEIDGSLTLDGVVGAARKGRVRYXDSDEIRKXIKDKSVEFLRKQLXNSVYGV 120
Db      55  P-TDSTLELDGSLNLDGVVGAARKGRVRYXDSDEIRKXIKDKSVEFLRKQLXNSVYGV 113
```

Mon Sep 13 10:31:15 2004

us-09-939-408a-21.rapb

Page 8

QY	12	TGCGGSAADTTEADIAISLQKALLLEHOLCGVLPTPSXSPFLGGGLGNSLPLEVVRGAMTRV	180
Dd	114	TGCGGSAADTTEADIAISLQKALLLEHOLCGVLPSPSPFLGGLGNSLPLEVVRGAMTRV	173
QY	181	NSLTRGHSAARLVLEALTNFLNHGITPVPVPRGITSASGDLSPSLYIAAITHGPDSKV	240
Dd	174	NSLTRGHSAARLVLEALTNFLNHGITPVPVPRGITSASGDLSPSLYIAAISHPDSKV	233
QY	241	HYHGEKXKMAAREALALGLPEPVVLGCEGLGVNGTAVSASVAITLADHAMLSLS	300
Dd	234	HYHGEKXKILVAREMALFNLEPVVLGCEKELGLVNGTAVSASVAITLADHAMLSLS	293
QY	301	QALTLALVEAMVGHASFFPHLDVTRPHPTQIEVAARITRLLEGSXPAYHNEEYVKD	360
Dd	294	QSLAMTVEAMVGHASFFPHLDVTRPHPTQIEVAGNIRKLBSGRPAVHNEEYVKD	353
QY	361	DEGLRQDRYPLRTSPQWLGPLVSDMIAHAALVLSAEGSTTDNPLIDVENKXTHGKNF	420
Dd	354	DEGLRQDRYPLRTSPQWLGPLVSDLIHAALVLTIEAGSTTDNPLIDVENKXSHGKNF	413
QY	421	QASAVNTEKTRIALALIGKLNFTQLTETMLNAGNRGLPSCLAEDPDSLHYCKGLDIA	480
Dd	414	QAAAVNTEKTRIGLAIQIGKLNFTQLTETMLNAGNRGLPSCLAEDPDSLHYCKGLDIA	473
QY	481	AAAVTSELGLAPVYTHVQPAEMGNQANSLALISARTAEADNLSILLATHLYCVLQ	540
Dd	474	AAAVTSELGLAPVYTHVQPAEMNQAANSLALISARTESNDVLSILLATHLYCVLQ	533
QY	541	AVDLRAMEFEFKQEPKXXXLLXOHFGXXXVYXXXXXELXXKXXKXLXRLQNTSYDL	600
Dd	534	AIDLRAHFEFKQEPKPAIVSLIDQHPGSAHMGSLRRELVENKXKLARLQNTSYDL	593
QY	601	EPRAHDAFSXATVTVEKXSSXXXXXVSLAAVNANKVAIAEKALSTRXVXXKPKXAPS	660
Dd	594	VPRWHDAFSFAAGTVVEVLSS---TSLSLAAVNANKVAIAESALSTRQVRETFWSAAT	650
QY	661	SSPALXYLSRFTVYKSPFYREELGVKARGDVFGLQKQVTTIGSNVSRITYEALSGRINX	720
Dd	651	SSPALXYLSRFTQILVAFYREELGVKARGDVFGLQKQVTTIGSNVSKITYEALSGRINX	710
QY	721	LVKOLA 726	
Dd	711	LVKOLA 716	
RESULT 13			
US-10-374-366-22			
US-Sequence 22, Application US/10374366			
US-Publication No. US20040014085A1			
GENERAL INFORMATION:			
APPLICANT: Tang, Xiao-Song			
APPLICANT: Milano, Joseph			
TITLE OF INVENTION: METHOD FOR THE RECOMBINATION OF GENETIC ELEMENTS			
FILE REFERENCE: CL1794 US NA			
CURRENT APPLICATION NUMBER: US/10/374,366			
CURRENT FILING DATE: 2003-02-26			
PRIOR APPLICATION NUMBER: 60/360,279			
PRIOR FILING DATE: 2002-02-26			
NUMBER OF SEQ ID NOS: 203			
SOFTWARE: PatentIn version 3.2			
SEQ ID NO 22			
LENGTH: 716			
TYPE: PRT			
ORGANISM: Rhodospiridium glutinis			
US-10-374-366-22			

```

Query Match      89.6%; Score 3015; DB 15; Length 716;
Best Local Similarity 83.9%; Pred. No. 2.9e-297;
Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3

OY      1 MAPSLDSIATSKANGKANGXHAAXXXXXXXXAGSGXLPITYXXITOLDIVEXXLD 60
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Dd	MAPSLDISISFANGVASAKQAVNGAS-----TNLAVAGSHLPTQVTDVIEVKMLAA	54
Qy	61 PXDDXXELDGYSLLTGDDVGAARKRXYKXVDSSEIKXIKDKSEFPLRXQJXNSYGV	120
Dd	55 P-DTSTLELDGYSLNJDVVASAARKRPRVXVDSSEIKSIKDKSEFPLRSQJNSYGV	113
Qy	121 TGFSGSADRTEDAIISLOKALLENHOCGLVPTFSXDSFXLGRGLENSJPLEVVRGAMTIRV	180
Dd	114 TGFSGSADRTEDAIISLOKALLENHOCGLVLPSSPDSFPLGRGLENSJPLEVVRGAMTIRV	173
Qy	181 NSLTREGSAVRVAVLEALTNFLNHGITTPIVPLRGITISASGJLSPLSTYAAAITGHDPDSKV	240
Dd	174 NSLTREGSAVRVAVLEALTNFLNHGITTPIVPLRGITISASGJLSPLSTYAAAISGHDPKSV	233
Qy	241 HVXHEGEEKTMKXAREALIFGLEPVYLGREGJGLVNGTAVASAMATLALHDAMLGJLS	300
Dd	234 HVXHEGEEKTLVAREKMAFLNIEPVYLGREGJGLVNGTAVASAMATLALHDAMLGJLS	293
Qy	301 QALTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVANNIRCTLLGSKFAVHEEVEVKD	360
Dd	294 QSLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLLEGSRFAVHEEVEVKD	353
Qy	361 DEGLRDRDRPLTISQWMLGPIVSMIHAAVLTSLBAGOSTTNDPLLDVENKXTHHGNGF	420
Dd	354 DEGLRDRDRPLTISQWMLGPIVSDLIHAAVLTSLBAGOSTTNDPLLDVENKXTHHGNGF	413
Qy	421 QASAVXNMTKRTLALILGKLNFTQLTETMLNAGNRCGLPSCLAEDPDSLSYHCKGJIDA	480
Dd	414 QAAAVANTMEKRTLGLAQIGKLNFTQLTETMLNAGNRCGLPSCLAEDPDSLSYHCKGJIDA	473
Qy	481 AAAYTSELGLAPVUTTHQVPAEMNNOAVNSIALISARTAEADVLSILATLTLXYVLO	540
Dd	474 AAAYTSELGLAPVUTTHQVPAEMNNOAVNSIALISARTIESNDVLSILATLTLXYVLO	533
Qy	541 AVDLRAMEFEFKQFEPXXXXLXOHFGXXXXLXXXXXELKXKXKXKXLEJGTSYDL	600
Dd	534 AIDLRATEFEFKQFEPALVSLIDQHFGSAMTGSNLRLDELVEKXNKTILAKRLBGTNSYDL	593
Qy	601 EPRRHDAFSAAGTVEVEXLSXXXXXXVSLAAVNAKVAAXAKAISLTXVXXAFXWAPSS	660
Dd	594 VEPKHDAFSAAGTVEVEXLS-----TSLSLAAVNTMKVAAAASASISLRQVRETFWSAAT	650
Qy	661 SSPALXYLSPTRTVLVSFVREELGVKARBGVFLGKQCVETIGSNVSRITYEAIKSGRINXV	720
Dd	651 SSPALXYLSPTRTQIILYAFVREELGVKARBGVFLGKQCVETIGSNVSKITYEALISGRINNV	710
Qy	721 LVTKMLA 726	
Dd	711 LITKMLA 716	

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

LOCATION: (502) . (502)
OTHER INFORMATION: X= Gly, Ala, Ser, Thr, Pro
NAME/KEY: UNSURE
LOCATION: (540) . (540)
OTHER INFORMATION: X= Thr, Ala, Ser, Pro, Gly
US-09-765-873A-38

Query Match 89.6%; Score 3014; DB 9; Length 716;
Best Local Similarity 83.9%; Pred. No. 3,7e-297;

Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

```
QY 1 MAPSLDLSIATSKXANGXNGXHAAXXASAXXXXXXAXAGSXLPTTXXTQLDIVEXKLAD 60
DB 1 MAPSLDLSIATSKXANGXNGXHAAXXASAXXXXXXAXAGSXLPTTXXTQLDIVEXKLAA 54
QY 61 PXTDDXELDGSYSLTLDGVVGAARKGRVYKDSDEIRXKIDKSVFELRQOLXSVYGV 120
DB 55 P-TDSTLELDGSLNLGDVVSAAKRGKRVYKDSDEIRSKIDKSVFELRQOLXSVYGV 113
QY 121 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSPFSPLGSGLENSLPLEVVRGAMTIRV 180
DB 114 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSPFSPLGSGLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPVLPGTISASGDLSPISYTAATGHPDSKY 240
DB 174 NSLTRGSAVRLVLEALTNFLNHGITPVLPGTISASGDLSPISYTAATGHPDSKY 233
QY 241 HVYHGXKIKMAREALFGLPEVPLGPKGGLGVNGTAVSASMATLALHDAHMLSLLS 300
DB 234 HVYHGXKIKMAREALFGLPEVPLGPKGGLGVNGTAVSASMATLALHDAHMLSLLS 293
QY 301 QALTALTYEAMVGAHSGHPFLHDVTRPHPTQIEVANIRTLGSGXFAYHHEEVYKXD 360
DB 294 QSLTAMTYEAMVGAHSGHPFLHDVTRPHPTQIEVANIRTLGSGXFAYHHEEVYKXD 353
QY 361 DEGIARODRYPRLTSPQWLGPLVSDMIAHAVALSLEAGOSTTNDPLLDVENKXTHHGGNF 420
DB 354 DEGIARODRYPRLTSPQWLGPLVSDMIAHAVALSLEAGOSTTNDPLLDVENKXTHHGGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 480
DB 414 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 473
QY 481 AAAYTSELGHLANPYTTHVOPAEWNGQAVNSIALISARPTAEANDVLSLLATHLVCVLQ 540
DB 474 AAAYTSELGHLANPYTTHVOPAEWNGQAVNSIALISARPTAEANDVLSLLATHLVCVLQ 533
QY 541 AYDLRAMEFEFKQFXPXXXXLLXQHFGXXXTXXXXXXELXXXKXKXRLLEQTNISYDL 600
DB 534 AYDLRAMEFEFKQFXPXXXXLLXQHFGXXXTXXXXXXELXXXKXKXRLLEQTNISYDL 593
QY 601 EPRMHDASFATGTVVEVLSXXXXXVSLAAVNAWKVAXAEKASLSTRYXXFXFWXAPSS 660
DB 594 EPRMHDASFATGTVVEVLSXXXXXVSLAAVNAWKVAXAEKASLSTRYXXFXFWXAPSS 650
QY 661 SSPALXYLSPRRVLVYSFVREELGVKARGDVFLGKQEVTTGSVNSRIYEAISGRINXV 720
DB 651 SSPALXYLSPRRVLVYSFVREELGVKARGDVFLGKQEVTTGSVNSRIYEAISGRINXV 710
QY 721 LVYQMLA 726
DB 711 LKMLA 716
```

RESULT 15
US-10-138-970A-22
Sequence 22, Application US/10138970A
Publication No. US20030079255A1

GENERAL INFORMATION:
APPLICANT: Sataelani, Sima
APPLICANT: Tang, Xiao-Song
APPLICANT: Qi, Wei Wei
TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

TITLE OF INVENTION: hydroxycinnamic Acid
FILE REFERENCE: CL1777
CURRENT APPLICATION NUMBER: US/10/138, 970A
CURRENT FILING DATE: 2002-07-23
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Microsoft Office 97
SEQ ID NO 22

LENGTH: 716

TYPE: PRT

ORGANISM: Rhodotocula glutinis

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (502) . (502)

OTHER INFORMATION: X=Gly, Ala, Ser, Thr, Pro

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (540) . (540)

OTHER INFORMATION: X=Thr, Ala, Ser, Pro, Gly

FEATURE:

NAME/KEY: MISC FEATURE

LOCATION: (540) . (540)

OTHER INFORMATION: Xaa=Thr, Ala, Ser, Pro, Gly

US-10-138-970A-22

Query Match 89.6%; Score 3014; DB 14; Length 716;
Best Local Similarity 83.9%; Pred. No. 3,7e-297;

Matches 609; Conservative 24; Mismatches 83; Indels 10; Gaps 3;

```
QY 1 MAPSLDLSIATSKXANGXNGXHAAXXASAXXXXXXAXAGSXLPTTXXTQLDIVEXKLAD 60
DB 1 MAPSLDLSIATSKXANGXNGXHAAXXASAXXXXXXAXAGSXLPTTXXTQLDIVEXKLAA 54
QY 61 PXTDDXELDGSYSLTLDGVVGAARKGRVYKDSDEIRXKIDKSVFELRQOLXSVYGV 120
DB 55 P-TDSTLELDGSLNLGDVVSAAKRGKRVYKDSDEIRSKIDKSVFELRQOLXSVYGV 113
QY 121 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSPFSPLGSGLENSLPLEVVRGAMTIRV 180
DB 114 TGFSGSADRTIEDAISLOKALLEHQLCGVLPSSPFSPLGSGLENSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPVLPGTISASGDLSPISYTAATGHPDSKY 240
DB 174 NSLTRGSAVRLVLEALTNFLNHGITPVLPGTISASGDLSPISYTAATGHPDSKY 233
QY 241 HVYHGXKIKMAREALFGLPEVPLGPKGGLGVNGTAVSASMATLALHDAHMLSLLS 300
DB 234 HVYHGXKIKMAREALFGLPEVPLGPKGGLGVNGTAVSASMATLALHDAHMLSLLS 293
QY 301 QALTALTYEAMVGAHSGHPFLHDVTRPHPTQIEVANIRTLGSGXFAYHHEEVYKXD 360
DB 294 QSLTAMTYEAMVGAHSGHPFLHDVTRPHPTQIEVANIRTLGSGXFAYHHEEVYKXD 353
QY 361 DEGIARODRYPRLTSPQWLGPLVSDMIAHAVALSLEAGOSTTNDPLLDVENKXTHHGGNF 420
DB 354 DEGIARODRYPRLTSPQWLGPLVSDMIAHAVALSLEAGOSTTNDPLLDVENKXTHHGGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 480
DB 414 QASAVXNTEKTRIALALIGKLNFTQLTMLNAGNNGRLPSCLAEDPSSLSYHCKGIDIA 473
QY 481 AAAYTSELGHLANPYTTHVOPAEWNGQAVNSIALISARPTAEANDVLSLLATHLVCVLQ 540
DB 474 AAAYTSELGHLANPYTTHVOPAEWNGQAVNSIALISARPTAEANDVLSLLATHLVCVLQ 533
QY 541 AYDLRAMEFEFKQFXPXXXXLLXQHFGXXXTXXXXXXELXXXKXKXRLLEQTNISYDL 600
DB 534 AYDLRAMEFEFKQFXPXXXXLLXQHFGXXXTXXXXXXELXXXKXKXRLLEQTNISYDL 593
QY 601 EPRMHDASFATGTVVEVLSXXXXXVSLAAVNAWKVAXAEKASLSTRYXXFXFWXAPSS 660
DB 594 EPRMHDASFATGTVVEVLSXXXXXVSLAAVNAWKVAXAEKASLSTRYXXFXFWXAPSS 650
QY 661 SSPALXYLSPRRVLVYSFVREELGVKARGDVFLGKQEVTTGSVNSRIYEAISGRINXV 720
DB 651 SSPALXYLSPRRVLVYSFVREELGVKARGDVFLGKQEVTTGSVNSRIYEAISGRINXV 710
```

Mon Sep 13 10:31:15 2004

us-09-939-408a-21.rapb

Page 10

Db 651 SSPALSTYLSPTOLIAFAFEELGVAKRGDVFLLGKQZVTIGSNVSKITYEAIKSGRINNV 710

OY 721 LVRKLA 726

Db 711 LVRKLA 716

Search completed: September 9, 2004, 10:24:27
Job time : 96.3859 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:50:23 / Search time 16.5685 Seconds
(without alignments)
2281.620 Million cell updates/sec

Title: US-09-939-408a-21

Perfect score: 3365
Sequence: 1 MAPSLDLSITSGXANGXNGX.....RIYFAIRGSRINVLVLMMA 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3021	89.8	1 P15544	rhodospirid
2	2922.5	86.8	1 P10248	rhodotorula
3	1294.5	38.5	1 O93967	amanita mus
4	1205	35.8	1 P45724	trifolium s
5	965	28.7	1 P35510	arabidopsis
6	958.5	28.5	1 P45729	petroselinu
7	957	28.4	1 P53777	pinus taeda
8	954	28.3	1 P27990	medicago sa
9	952.5	28.3	1 P17117	oryza sativ
10	949	28.2	1 O64963	prunus aviu
11	947	28.1	1 P45734	trifolium s
12	946.5	28.1	1 P45728	petroselinu
13	946	28.1	1 P27991	glycine max
14	943.5	28.0	1 Q96845	arabidopsis
15	943	27.9	1 P25872	nicotiana t
16	940	27.9	1 O23924	digitalis 1
17	939.5	27.9	1 P45724	arabidopsis
18	938	27.9	1 P35511	lycopersico
19	933.5	27.7	1 Q01861	psium sativ
20	932	27.7	1 Q01861	psium sativ
21	931.5	27.7	1 Q01861	psium sativ
22	930.5	27.7	1 Q01861	psium sativ
23	929.5	27.6	1 Q01861	psium sativ
24	929	27.6	1 Q01861	psium sativ
25	929	27.6	1 Q01861	psium sativ
26	929	27.6	1 Q01861	psium sativ
27	928.5	27.6	1 Q01861	psium sativ
28	927.5	27.5	1 Q01861	psium sativ
29	927	27.5	1 Q01861	psium sativ
30	926.5	27.5	1 Q01861	psium sativ
31	925	27.5	1 Q01861	psium sativ
32	924.5	27.5	1 Q01861	psium sativ
33	924.5	27.5	1 Q01861	psium sativ

ALIGNMENTS

RESULT 1

1 P15544; STANDARD; PRT; 716 AA.

09sm9 cicer ariet
O49835 lithospermu
E31426 solanum tub
Q42667 citrus limo
E31425 solanum tub
P45730 populus tri
P45725 arabidopsis
P19143 phaseolus v
Q43052 populus kit
P45731 populus kit
P41456 ipomoea bat
Q43210 triticum ae

AC P15544; 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL.
OS Rhodospiridium toruloides (Yeast) (Rhodotorula gracilis).
OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
OC Microbotryomycetidae; Sporidiobolales; Rhodospiridium.
OX NCBI_TaxID=5286;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CBS 14;
RX MEDLINE=92227061; PubMed=1773059;
RA Rasmussen O.F., Orum H.;
RT "Analysis of the gene for phenylalanine ammonia-lyase from
Rhodospiridium toruloides";
RL DNA Seq. 1:207-211(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=IFO 0559;
RX MEDLINE=88112870; PubMed=2828184;
RA Anson U.G., Gilbert H.U., Orum J.D., Minton N.P.;
RT "Complete nucleotide sequence of the Rhodospiridium toruloides gene
coding for phenylalanine ammonia-lyase";
RL Gene 58:189-199(1987).
CC -|- FUNCTION: Catalyzes the nonoxidative deamination of L-
phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC -|- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
CC -|- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -|- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
which is formed autocatalytically by cyclization and dehydration
of residue Ala-Ser-Gly (By similarity).
CC -|- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
DR EMBL; X51513; CAA35886.1; -;
DR EMBL; M18261; AAA33883.1; -;
DR EMBL; X12702; CAA31209.1; -;
DR PIR; A29607; A29607.
DR PIR; A56628; A56628.
DR HSP; P21310; I88F.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3-lyase.
DR InterPro; IPR005922; Phe_am_lyase.

DR Pfam; PF00221; PAL; 1.
 DR TIGRfams; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KM Lyase; Phenylpropanoid metabolism.
 FT CROSSLINK 211 213 5-imidazolinone (Ala-Gly)
 FT (By similarity).
 FT MOD_RES 212 212 DH4 (2,3-DIHYDROALANINE)
 FT (BY SIMILARITY).
 FT CONFLICT 4 37 SLDSISHSFANGVSAKQAVNGASTLVAAGSHL ->
 FT RPTSGQARPC (IN REF. 2).
 SQ SEQUENCE 716 AA; 76879 MW; 0CIDF6176944B5E6 CRC64;

Query Match 89.8%; Score 3021; DB 1; Length 716;
 Best Local Similarity 84.0%; Pred. No. 9, 8e-227;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAPSLDSTATSXANGXXGXAAXSXXXXXXXXXAGSKLPTTXXTQDLIVEXXLAD 60
 DB 1 MAPSLDSTATSXANGVSAKQAVNGAS-----TLVAAGSHLPTTQVTVQVIVKMLAA 54

QY 61 PXTDDXXELDGYSLTLDVVGAAKRGXVYVXDSEIRKXIDKSYEFLRXQIXNSVGYT 120
 DB 55 P-TDSTLELDGYSLTLDVVGAAKRGXVYVXDSEIRKXIDKSYEFLRXQIXNSVGYT 113

QY 121 TGRGSGADTTEDAISIQKALLHQCGLVPSXDSFKLGRGLNSLPLEVVRGAMTIRV 180
 DB 114 TGRGSGADTTEDAISIQKALLHQCGLVPSXDSFKLGRGLNSLPLEVVRGAMTIRV 173

QY 181 NSLTRGHSARLVLENTNFNLHGKITPIVPRGTSSASGDSPLSYTAAATGHPDSKV 240
 DB 174 NSLTRGHSARLVLENTNFNLHGKITPIVPRGTSSASGDSPLSYTAAATGHPDSKV 233

QY 241 HVHHEGKETKXAREALFGLSEPVVLGREGSLVNGTAVSASMAATLADAMLSLS 300
 DB 234 HVHHEGKETLVAREMALFNLSEPVVLGREGSLVNGTAVSASMAATLADAMLSLS 293

QY 301 QALTALTEAMVGHASFPFLHDVTRPHPTQIEVARNRITLSEGSRAVHHEEVKVD 360
 DB 294 QSLTAMTEAMVGHASFPFLHDVTRPHPTQIEVARNRITLSEGSRAVHHEEVKVD 353

QY 361 DESILRQDRYPLRTSPQWLGPVSDMIAHAVALSLDAQSTTDNPLIDVENKXTHGDNF 420
 DB 354 DESILRQDRYPLRTSPQWLGPVSDMIAHAVALSLDAQSTTDNPLIDVENKXTHGDNF 413

QY 421 QAAVYNTMEKTRFLALATGKNTFQLTMLNAGMNRGLPSGLAABDSLSYHCKGLDIA 480
 DB 414 QAAVYNTMEKTRFLALATGKNTFQLTMLNAGMNRGLPSGLAABDSLSYHCKGLDIA 473

QY 481 AAAYTSELGLANPVTTHVQPAEMGNOAVNSIALISASRTAEANDVLSILATLHYCQLQ 540
 DB 474 AAAYTSELGLANPVTTHVQPAEMGNOAVNSIALISASRTAEANDVLSILATLHYCQLQ 533

QY 541 AVULRAMEFEFKQEPKXXXLLXQFGKXXTXXXXXXELKXKVKXILKRLLEQTNLYD 600
 DB 534 AIDLRRIEIEFKQEPKXXXLLXQFGKXXTXXXXXXELKXKVKXILKRLLEQTNLYD 593

QY 601 EPRMHDASFATGTVEEXLSXXXXXSVSLAAVNAKVAAXAKALSLTREVXXXQWXPSS 660
 DB 594 VPRMHDASFATGTVEEXLS----TSLSLAAVNAKVAAXAKALSLTREVXXXQWXPSS 650

QY 661 SSPALXYLSPTRTVLVSFYREELGVYKARGDVLGKQVTTGNSVSRITYEAKSGRLNV 720
 DB 651 SSPALXYLSPTRTVLVSFYREELGVYKARGDVLGKQVTTGNSVSRITYEAKSGRLNV 710

QY 721 LVKMLA 726
 DB 711 LVKMLA 716

RESULT 2
 PALLY_RHORB STANDARD; PRT; 713 AA.
 AC P10256;

DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PAL.
 OS Rhodotorula rubra (Yeast) (Rhodotorula mucilaginosa).
 OC Eukaryota; Fungi; Basidiomycota; Uredinomycetes;
 OC Microbotryomycetidae; Sporidiobolales; mitosporic Sporidiobolales;
 OC Rhodotorula.
 OX NCBI_TaxId=5537;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL Y-15597;
 RX MEDLINE=89083582; PubMed=3205749;
 RA Filipula D., Strausberg R.L., Vastle C.A., Levy A.;
 RT "Nucleotide sequence of gene for phenylalanine ammonia-lyase from
 Rhodotorula rubra.";
 RL Nucleic Acids Res. 16:11381-11381(1988).
 RN (2)
 RP DETAILS.
 RC STRAIN=NRRL Y-15597;
 RX MEDLINE=89083582; PubMed=3205750;
 RA Filipula D., Strausberg R.L., Vastle C.A., Levy A.;
 RT "cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes
 reveal genomic intron deletions.";
 RL Nucleic Acids Res. 16:11382-11382(1988).
 CC -1- FUNCTION: Catalyzes the nonoxidative deamination of L-
 phenylalanine to form trans-cinnamic acid and a free ammonium ion.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 which is formed autocatalytically by cyclization and dehydration
 of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL: X13094; CAA31486.1; -.
 CC PIR: S01999; S01999.
 CC HSSP: P21310; 1B8F.
 DR InterPro: IPR008948; L-Asparaginase-like.
 DR InterPro: IPR001106; Phe/His NH3-lyase.
 DR InterPro: IPR005922; Phe_am_lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRfams: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KM Lyase; Phenylpropanoid metabolism.
 FT CROSSLINK 217 219 5-imidazolinone (Ala-Gly)
 FT (By similarity).
 FT MOD_RES 218 218 DH4 (2,3-DIHYDROALANINE)
 FT (BY SIMILARITY).
 SQ SEQUENCE 713 AA; 76000 MW; 6EB8317CBB037DB8 CRC64;

Query Match 86.8%; Score 2922.5; DB 1; Length 713;
 Best Local Similarity 82.0%; Pred. No. 4, 3e-219;
 Matches 595; Conservative 33; Mismatches 85; Indels 13; Gaps 5;

QY 1 MAPSLDSTATSXANGXXGXAAXSXXXXXXXXXAGSKLPTTXXTQDLIVEXXLAD 60
 DB 1 MAPSLDSTATSXANGVSAKQAVNGAS-----TLVAAGSHLPTTQVTVQVIVKMLAA 59

QY 61 PXTDDXXELDGYSLTLDVVGAAKRGXVYVXDSEIRKXIDKSYEFLRXQIXNSVGYT 120
 DB 60 AGATDQITLSDGYSLTLDVVGAAKRGXVYVXDSEIRKXIDKSYEFLRXQIXNSVGYT 119

QY 121 TGRGSGADTTEDAISIQKALLHQCGLVPSXDSFKLGRGLNSLPLEVVRGAMTIRV 180

```

Db      120 TGGGAGDADRTRDALSLOKALLBHOLOCVLPRTSMGCFALGGGLNSLPLEVRKAMTRV 179
QY      181 NSITRGSAVRLVLPALTNPLNHGITPVPVIRGTTISAGSLSPSYIAAITHPDSKY 240
Db      180 NSITRGSHSVRLVLPALTNPLNHGITPVPVIRGTTISAGSLSPSYIAAITHPDSKY 239
QY      241 HYHXEKXKIMKAREALFGLPEVPLGPKGELGWNSTAVASAMATLADHMLSLLS 300
Db      240 HV--DG--KIMAGQERIALKGLQPVVLGPKGELGWNSTAVASAMATLADHMLSLLS 295
QY      301 QALTALTVEAMVGHAGSFHPPLHDVTRPHPTQIEVARNIRTLLEGSKFVHNEEVKXD 360
Db      296 QALTALTVEAMVGHAGSFHPPLHDVTRPHPTQIEVARNIRTLLEGSKFVHNEEVKXD 355
QY      361 DEGIILRODVPYLRTPBOWLGPLVSMIHAHVLSTLGAQSTTDNPLLDLENKMHGQAF 420
Db      356 DEGIILRODVPYLRTPBOWLGPLVSMIHAHVLSTLGAQSTTDNPLLDLENKMHGQAF 415
QY      421 QASAVXNTEKTRIALALIGKLNFTQITEMLNAGKNGRLPSCLAEDPSLSYHCKGLDIA 480
Db      416 MASSVGNTEKTRIALVALMGKVSFTQITEMLNAGKNGRLPSCLAEDPSLSYHCKGLDIA 475
QY      481 AAAYTSELGHANPVTVTHVQPAEMGNQAVNSIALISARRTAEANDVLSLLATLYCVLQ 540
Db      476 AAAYTSELGHANPVTVTHVQPAEMGNQAVNSIALISARRTAEANDVLSLLATLYCVLQ 535
QY      541 AVDLRAMEFEFKQFPKXXLLXQHPGXXXTXXXXXXELKXXVXXKXRLBETNSYDL 600
Db      536 AVDLRAMEFEFKQFPKXXLLXQHPGXXXTXXXXXXELKXXVXXKXRLBETNSYDL 590
QY      601 EPRWDHAFSXAATGVVEXLSXXXXXVSLAAVNAKVAKAKAISLTXXVXXKXWAPSS 660
Db      591 EGRWHDTFVAVGAYVEALAG--GEVSLASLNAKVAKAKALALTRVSDSWAPSS 647
QY      661 SSPALXYLSPRRVLYSPFREELGYKARGVFLGQEVTTIGSNVSRIYELKSGRLNXY 720
Db      648 SSPALXYLSPRRVLYSPFREELGYKARGVFLGQEVTTIGSNVSRIYELKSGRLNXY 707
QY      721 LVKMLA 726
Db      708 LVKMLA 713

RESULT 3
ID      PALY AMAMU          STANDARD:          PRT:          740 AA.
AC      093967;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DE      10-OCT-2003 (Rel. 42, Last annotation update)
DN      Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN      PAL.
OS      Amanita muscaria (Fly agaric).
OC      Agaricales; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OX      Agaricales; Amanitaceae; Amanita.
RN      NCBI_TaxID=41956;
[1]
SEQUENCE FROM N.A.
RP      Nehls U., Mikolajewski S., Eicke M., Hamp R.;
RT      'Sugar and nitrogen-dependent regulation of an Amanita muscaria
RL      phenylalanine ammonium lyase gene.';
RT      Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
CC      - FUNCTION: Catalyzes the nonoxidative deamination of L-
CC      phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC      - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC      - PATHWAY: Phenylpropanoid biosynthesis; first step.
CC      - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC      - PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC      which is formed autocatalytically by cyclization and dehydration
CC      of residues Ser-Ser-Gly (By similarity).
CC      - SIMILARITY: Belongs to the PAL / histidase family.

```

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AJ010143; CA009013.1; -.
DR      HSSP: P21310; 188F.
DR      InterPro: IPR008948; L-Asparagine-lyase.
DR      InterPro: IPR001106; Phe/His NH3lyase.
DR      InterPro: IPR005922; Phe_am_1lyase.
DR      Pfam: PF00221; PAL; 1.
DR      TIGRPFAM: TIGR01226; phe_am_1lyase; 1.
DR      PROSITE: PS00488; PAL_HISTIDASE; 1.
KW      Lyase; Phenylpropanoid metabolism.
FT      CROSSLINK 224 226 5-imidazole-one (Ser-Gly)
FT      MOD_RES 225 225 DHA (2,3-DIHYDROQUANTANINE)
FT      FT Best Local Similarity 42.6%; Pred. No. 1,1e-92;
SQ      SEQUENCE 740 AA; 80156 MW; B055CF3D8B7BCEB CRC64;
Query Match 38.5%; Score 1294.5; DB 1; Length 740;
Best Local Similarity 42.6%; Pred. No. 1,1e-92;
Matches 288; Conservative 123; Mismatches 240; Indels 25; Gaps 9;
QY      68 ELDGSLTLGADVAGARKGRKRVYVDSDEIRKXIDKSYEFRLXQUN--SYGVYTFEGG 125
Db      69 KVDGGLTLAAVAAARNAVAELDESPVKEVRKSLAIAMKVSFGASYGISTFGG 128
QY      126 SADRTEDALSLQKALLBHOLOCVLPSTX--DSFXLGRGLNSLPLEVRKAMTRVNS 182
Db      129 SADRTDXKMLGFPALIQHVGILPTSTPEPLDVLPIQDANNSTMPAMIRGAILIRNS 188
QY      183 LTRGSAVRLVLPALTNPLNHGITPVPVIRGTTISAGSLSPSYIAAITHGDSKYH- 241
Db      189 LTRGSHGIRWELIEKREBLANVLPVPLRGSISSGDSLSPSYIAGTTIGNSIKVYH 248
QY      242 -VHXHEKXKIMKAREALFGLPEVPLGPKGELGVNCTAVASAMATLADHMLSLLS 300
Db      249 GPKSGIRQIGSSKVALAHNIEPPLESKEPLGINTASASVALALNEALHIVLLA 308
QY      301 QALTALTVEAMVGHAGSFHPPLHDVTRPHPTQIEVARNIRTLLEGSKFVHNEEVKXD 360
Db      309 QVCTAMGTEALIGTASHAPFIHATARPPOQVCAEINWILDGSKLQLEHEEVLIED 368
QY      361 DEGIILRODVPYLRTPBOWLGPLVSMIHAHVLSTLGAQ--STDNPLLDVENKXTHRGN 419
Db      369 DXYTRRODVPYLRTPBOWLGPLVSMIHAHVLSTLGAQ--STDNPLLDVENKXTHRGN 428
QY      420 FOASAVXNTEKTRIALALIGKLNFTQITEMLNAGKNGRLPSCLAEDPSLSYHCKGLDI 479
Db      429 FOAMAVTNAMKTRIALAHVGLLFQSTELVNPAMRGLPPSVAAIDPSPINAKGLDI 488
QY      480 AAAYTSELGHANPVTVTHVQPAEMGNQAVNSIALISARRTAEANDVLSLLATLYCVL 539
Db      489 ATPAAVAAE---ATGCPHTIQSAEKNQAVNSIALISARRTAEANDVLSLLATLYCVL 544
QY      540 QAVDLRAMEFEFKQFPKXXLLXQHPGXXXTXXXXXXELKXXVXXKXRLBETNSYD 599
Db      545 QAVDLRALQREFLPGLDIIIEELRSSFGFLS--SEQEKIQLQNLTAFAEHDHDKTTMD 603
QY      600 LEPRHDAFSXAATGVVEXLSXXXXXVS-----LAAVNAKVAKAKAISLTXXVXXF 654
Db      604 NTDRMTTMAATSSSVLTFPFDGSGAVPSSCDLLSSVSFQSSVATRRSSVLTMDLKEY 663
QY      655 WAPSSSPALXYLSPRRVLYSPFREELGYKARGVFLGQEVTTIGSNVSRIYELKSGRLNXY 708
Db      664 TRGDGEPYLPASQYIG-KTRPYQIFIRTTIGVRKGSSEYKVFYNGLGEVDVTTIGQNSRI 722
QY      709 YEAIKSGRLNXYLVLM 724

```


Db 723 YESIRDCGQSIIVSL 738

RESULT 4
ID PALT USTWA STANDARD; PRT; 724 AA.

AC 096V77;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PALI;
OS Ustilago maydis (Smut fungus).
CC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
CC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim S.-H., Macdonald K., Virmant D., Wake K., Kronstad J.W.,
RT "Cloning and disruption of a phenylalanine ammonia-lyase gene from
RT Ustilago maydis";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the nonoxidative deamination of L-
CC phenylalanine to form trans-cinnamic acid and a free ammonium ion.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidine family.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----
DR EMBL; AF306551; AL009388.1;
DR InterPro; IPR008948; L-Asparatase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TricRfam; TRC001226; Phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism.
FT CROSSLINK 205
FT 5-imidazolone (Ala-Gly)
FT (By similarity).
FT MOD_RES 206 206 DHA (2,3-DIDEHYDROALANINE)
FT (BY SIMILARITY).
SQ SEQUENCE 724 AA; 79277 MW; 02A73C97C34C01AB CRC64;

Query Match 35.8%; Score 1205; DB 1; Length 724;
Best Local Similarity 41.6%; Pred. No. 9, 1e-86;
Matches 287; Conservative 114; Mismatches 245; Indels 44; Gaps 15;

QY 69 LDGYSLTLDGVVGAARKGRVXVXDSEIRXKIDSVFELKXQV--XNSVYGVTTGGGS 126
DB 41 IDGNLKLQGLVNSAAYGVHTPRPSAETKRKIDSVQSLAKLDGGSITGIVTGGGS 100
QY 127 ADRTEDATSLQKALHEQLCGVLP-----SXDSFXLG-RGLENSD-PLEVVRG 174
DB 101 ADSRTANTRALQALHMOCCGVLVPSTFTGEBSSAPFALPLDTRESSLIMEAVVRG 160
QY 175 AMTIRNNSLTREGSNAVRLVLELNT-FLNHGITPTVPLRGITNSGSLSLSTIAAIT 233
DB 161 AIVVRSLSSIMRGHSGVWEVLDKMQKLFQNNVTVVVPRSSISGSLSPSLVAAGALA 220
QY 234 GHDSKVHYHGEKEXKI-MKAREALALFGLPEPVVGLPKEGLVNGVTAVSAMTALHD 292

Db 221 GQRGIYCTVTDGRGRVVTADACRMAKITPVQYEPKEALGLNGTAFSAVGLATYE 280

QY 293 AMLSLSGALPTALTEAVNGHAGSFHPLHDVTRPHPTQIEVANRTITLEGXFVAVH- 351
DB 281 AENLASLTQLTAAVVALKGTDSFAFPIHEIARPHGQIKSAKFIKRAHLSGRLEHL 340
QY 352 -HEEVKVKDDEGITRORRYPLRTSPQWGLGVSDMKHAHAVSLSEAGSTTNPLIDV- 409
DB 341 ENEKVLTSEBDGTLRQRYTLQTSQWVGGLDIENAKSVDFEI-NSTDPMIDPY 399
QY 410 -ENKTHHGNGFQASAVNVTWETKRLALALGKINFTQTEMLNAGNRGIPSCIA-ED 467
DB 400 DDDGRIHHGNGFQAAVMNNAVEKIRLALCAAGKMTFQGMTELVNPNRRGLPNTLSTPD 459
QY 468 PSLSTHCKGLDIAAAAYSEIHLANPYTHVQPAENGQAUNSLAISARPAEANDVL 527
DB 460 LSLNPAKGINITALSVTSELMFLGNPVSTHVQSAEVANQAFNSLALISGRQTLQATECL 519
QY 528 SLLLAHTLYCVLAAYDLAMEPEFKQEPXXXXLLKHQFGXXXXXXELXXXXKXX 587
DB 520 SWIQWLSYLLQALDIALQYKABQPLTILASLSHRS-EMDETQKQETIAQVLS 578
QY 588 LKRLLEQNSYDLBPRMHDASFATGVVEXLSS-XXXXXVSLAANAKVAXAEKALS 645
DB 579 MGRKLDSTSSKDLRDLVETVQDASSVLYRFSELPGCGADPLRNIVKRAVGVADTEK 638
QY 646 LTRXYRXXFWKAPSSSPALKYLSPRTRVLYSFVREELGYKARRGDVFLEKQSV----- 699
DB 639 ITRQVTEIEDNPFVCHAS--HLGKTRGAEEVFKTLGVPMH-----GREENLNEPKGE 690
QY 700 -----TIGSNVRIYEAIKSGRINXVLVKM 724
DB 691 FEQWNTTGGYVSVIAYISIRDEGLVNMKSEL 720

RESULT 5
ID PALT ARATH STANDARD; PRT; 725 AA.

AC P3551C; Q92QD6;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase I (EC 4.3.1.5).
GN PALI OR AT2G37040 OR T1J8.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid I; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=95195160; PubMed=788622;
RA Wanner L.A., Li G., Ware D., Somsich I.E., Davis K.R.,
RT "The phenylalanine ammonia-lyase gene family in Arabidopsis
RT thaliana";
RL Plant Mol. Biol. 27:327-338 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.W., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Bueli C.R., Ketchum K.A., Lee J.-J., Romling C.M., Xoo H.L.,
RA Moffet K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Frazer C.M.,
RA Venter J.C.,
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana";
RL Nature 402:761-768 (1999).
RN [3]

DR PIR; S48726; S48726.
 DR HSBP; P21310; 188F.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMs; TIGR01226; Phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 204 206
 (By similarity).
 FT MOD_RES 205 205
 DHA (2,3-DIHYDROALANINE)
 (By similarity).
 SQ SEQUENCE 718 AA; 78165 MM; 17786451PFDEBC35 CRC64;

Query Match 28.5%; Score 958.5; DB 1; Length 718;
 Best Local Similarity 35.2%; Pred. No. 1.2e-66;
 Matches 264; Conservative 115; Mismatches 285; Indels 87; Gaps 21;

14 NCXXNKGHAXXAXXXXXXAXXAGSLPTTXXTOLDIYEXXIADPTDXXELDGY 73
 5 NGTTNG-HANGNGLDLCKRDEPLMWGVAEALTGSHLDEVKRYVAE-YRKPVYKLGSET 62
 74 LTLGDVVG-AARKGRYRVVDSDETRKIDKSEVFELRXQLXNSV--YGVTTGFGSGADTR 130
 63 LTIQVVAISARDSDGVKELSEBARAGYKASDMVMSMKKGDSDYVTTGFGATSHR 122
 131 TEDAISOQALLHQLGCVLPTSGXDSFKLGRGLE--NSLPLEVVGAMTIRVNSLTRGH 187
 123 TKOGALQKLELLEFLNAGI-----FGSGAEAKNNVTLPHSATRAMLVRIINTLQGY 173
 188 SAARLVVLEALTNFNLHGITPIVPLRGTTASAGDISPISTYAAATGHPDKVYHGX 247
 174 SGREFILPAITKFLMHNITPCLPRGTTASGDIVPASTYAGLTGEPNSKA-VGPTGV 232
 248 EKIMKAREALFGLER--VVLGPKGGLVNGTVASASMTALHDHMLSLSQALTA 305
 233 --TLSPBEAFKLAVGEGFELQPKREGALVNGTVASGMSMTLFEANILAVLEWMSA 290
 306 LTVANVGHAGSHHPLHDVTRPHPTQIEVANRITLLEGSFAVHEEYKVKDCEGIL 365
 291 IFAEVVOGKR-BETDHLTRKLGKHGQIEEALIMHIIIDGSAY-VKAQKHEMDPLQKP 348
 366 RODRYPLRTSPOMLGLVSDMTAHAVLSLEAGOSTTNPPLIDENKXTHHGNFQASAV 425
 349 KQRYVALRTSPOMLGPQI-EVIRSETKMERELNSVNDNPILIVSRKKAHIGNFQSP 407
 426 XNTEKTRIALALLIGKLNFTQLEMLNAGNRGLPSCLA-AEDPSLSYHCKGLDIAAAY 484
 408 GVSMDTRLAIAIGLMPAQSELVNDFYNNGLPSNLGSRNPSLDVGFAGABIAVASY 467
 485 TSEIGLANPVTHHOPAMGNQAVNSLITISARTAAADVLSLATHYCYLQAVDL 544
 468 CSBLOGLAPVTHHOSAEQHNQDVNSJGLISRTKTSAAVEILKLMSTTFVLGCOAIDL 527
 545 RAMEPEFKQFPYXXXXLXQHFQGXKTXXXXXXELKKXV-----XKXLX 589
 528 RHLEERLTK-----STVKNVTGVAVARVLTWNGVNGELHPSRCECDL 569
 590 KALEQ-----TNSYDLERPMHAFSAKGTVEVEXLSXXXXXVSLAANAMKVA 638
 570 RAVDREYTFAYIDPQASATYPLMQCLRE-----TLVEHALNNGDKERNLSTISFOKIA 622
 639 XAEKATIS--LTXVXXXFWXAPSSSPAL--XVLSPTRVLYSFVREELGVKARAGDVEL 694
 623 AFEDELKALLPREVETA-RAALESQNPALPNRIKESGYPYKRYVREELGI-----EYLT 676
 695 GKQEVLTIGSNVSRITPAIKGRINXVLYKML 725
 677 GEKVRSPGEPEKFTAMSKGEIIDPLLECL 707

PALY_PINTA
 ID PALY_PINTA STANDARD; PRT; 754 AA.
 AC P52777;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN PAL.
 OS Pinus taeda (loblolly pine).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
 OX NCBI_Taxid=3352;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Needle;
 RA Zhang X.H., Chiang V.L.;
 RU Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC - CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC - PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC - PTM: Contains an active site 4-methyliden-indazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (by similarity).
 CC - SIMILARITY: Belongs to the PAL / histidase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U39792; AAA8489.1; -
 DR PIR; T09777; T09777.
 DR HSBP; P21310; 188F.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMs; TIGR01226; Phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism.
 FT CROSSLINK 206 208
 (By similarity).
 FT MOD_RES 207 207
 DHA (2,3-DIHYDROALANINE)
 (By similarity).
 FT SEQUENCE 754 AA; 82600 MM; 59897746AACCA398 CRC64;

Query Match 28.4%; Score 957; DB 1; Length 754;
 Best Local Similarity 36.9%; Pred. No. 1.7e-66;
 Matches 252; Conservative 109; Mismatches 258; Indels 64; Gaps 18;

69 LDGYSLTGVDVGAARKGRYRVVDSDEIKXIDKSEVFELRXQLX--NSYVGGTTTFGGS 126
 64 IEKSLTISDVAAARASQYKVKLDAAASRVEBSNNVLTQMTGTDYGTGTFGAT 123
 127 ADRTEDAISQKALLHQLGCVLPTSGXDSFKLGRLENSLPLEVVGAMTIRVNSLTNG 186
 124 SHRTNOCAGELQKLELLEFLNAGV-----LGKCPENVLSDDTTRAMLVRYTLLDG 174
 187 HSAVRLVLEALTNFNLHGITPIVPLRGTTASAGDISPISTYAAATGHPDKVYHXBHG 246
 175 YSGVRMDILEVVEKLANMLTPKPLRGITTTASGDIVPASTYAGLTGEPNSRVR-SRQD 233
 247 XEKIMKAREALFGLER--PVVLGPKGGLVNGTVASASMTALHDHMLSLSQALTA 305
 234 IE--MSGAEALKKVGLKPELQPKREGALVNGTVGALASIVCFDANVLLALSEVISA 291
 306 LTVANVGHAGSHHPLHDVTRPHPTQIEVANRITLLEGSXFAVH--HEEVVYKXD 361

RESULT 9
PAL1_ORYSA STANDARD; PRT; 701 AA.
ID PAL1_ORYSA
AC 164717
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
PAL.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eubacteriodes; Oryzae; Oryza.
OC NCBI_Taxid=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA MEDLINE=90032676; PubMed=2806257;
RX Mitani E., Ozeki Y., Matsuo M., Kozuka N., Tanaka Y.;
RT "Structure and some characterization of the gene for phenylalanine
ammonia-lyase from rice plants.";
RL Eur. J. Biochem. 185:19-25(1989).
CC
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
first reaction in the biosynthesis from L-phenylalanine of a wide
variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
which is formed autocatalytically by cyclization and dehydration
of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X16093; CA34226.1; -.
DR HSSP; P21310; 1B8F.
DR Gramene; P14717; -.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRfam; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Phenylpropanoid metabolism; Multigene family.
FT FT CROSSLINK 189 191 5-imidazolone (Ala-Gly)
FT FT MOD_RES 190 190 DHA (2,3-DIDEHYDROALANINE)
FT FT (By similarity).
FT FT (By similarity).
SQ SEQUENCE 701 AA; 75808 MW; 2BF9ADF0B8E04DE CEC64;
Query Match 28.3%; Score 952.5; DB 1; Length 701;
Best Local Similarity 35.9%; Pred. No. 3.5e-66;
Matches 254; Conservative 117; Mismatches 266; Indels 71; Gaps 21;
QY 40 GSXLPXTXXHOLDIVXXLADPXTDDXXELDQSYLTLDGV--VGAARKGRVRYXSDDEI 97
DB 16 GAATAAAGSHLDEVGAWAQ--FREPLVKTIGATIRVQVAAVAQAQAAVAVELEDEA 74
QY 98 RKKIDKSVFLRXOLAN--SVYGVTTFGSGADRTEDPAISLQVALLHEHQLCGVLPSTXD 155
DB 75 RPRVKAASWMLTCAHGGDIDYGVTTGGCTSHRTKGPALQVYELLRYLNAGIFGIGSD 134
QY 156 SPYLGGLGSLPLEVYRGAMTIRVNSLTGRGSAVRLVVLALNPLNHHGTPVPLRGT 215

DB 135 G-----HTLPSETRVAAMLVRINTLLQGYSGIREILLEATIKLTINTGTPCLPDRGT 186
QY 216 ISASGDLSPASTYIAALITGHDPDSKVHVHEKXKIMKAREALAFGLEP--VVLGREG 273
DB 187 ITASGDLVPLSTIAILLITGRPAAQ--ISPDRK--VDAAEFKLAGIEGGFTLNPEGL 243
QY 274 GLVNGTAVASAVATLADLADMLSLISQALTAITVEAMVAGSFFHFDVTRPPTQI 333
DB 244 AIVNGTSVGSALAAVMDANILAVLSVLSAVCEVMNGKPEYDHLTKL--KHPGSI 302
QY 334 EVARNIRTLBGSXAVAHHEEVKYKDEGL--RQRYPLRISPOLGVLSDVYHAHA 391
DB 303 DAAAMEHILAGSSFSMSHAK--KYNEMDPLLKQORVALRTSPQWLGPQI--QVIRA-A 357
QY 392 VLSLEAG--QSTDNLDIVENKXTHHGGNFPASAVXNTEKTRTALALIKLNFOTLEM 450
DB 358 TKSIREVNSVNDPVIDVHKGKALHGGNPGCTIGVSMNARLAINIKMFAQPSL 417
QY 451 INAGNNRGLPSCLA-AEDPSLSYHCKGLDIAAAVTSSELGLANPVTHVQPAEMGNQAV 509
DB 418 VNEFYNNGLTSLAGSRNPSIDYGFKEIIMASYSSELQYLANPITNHVQSAECHOVDV 477
QY 510 NSALISARRAENDVLSLLATHIVCVQAVDLRAMEEFKQFPPKXXLLXQHPGX 569
DB 478 NSGLVSARKTLEAVDILKMTSTYIYALCQAVDLRHEENIYSSVNCVTQVAKKVLTM 537
QY 570 XXXXXXXXELXXKVKXKXKLEQTNSDLEP--RWHDAPFSXATGVVE-XLSXXXXX 626
DB 538 NTGDLSSARPESEKULLAIR--EAVSYADDPESANVPLMQKLRALVVEHALTSGDRA 596
QY 627 VSLAIVNAWKYAKXAEKALSLTRVXXKFWAPSSSP-----ALYILSP 670
DB 597 RGLRLYQDHQYRG-----APLCAAGDGRRRRRRORTAPYANRIVES 640
QY 671 RTRVLYSFREELGVKARGDVPF--GROEVTIGSNVRIRIYEAISGRI 717
DB 641 RFPPIRYRREELG-----VFLTGRKLSPECECKVVLGISQKTL 682
RESULT 10
PAL1_TRISU STANDARD; PRT; 725 AA.
ID PAL1_TRISU
AC P45734;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PAL1.
OS Trifolium subterraneum (Subterranean clover).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
OX NCBI_Taxid=3900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Karriidae; TISSUE=leaf;
RA MEDLINE=94171061; PubMed=8125321;
RX Howles P.A., Arioli T., Weimann J.J.;
RT "Characterization of a phenylalanine ammonia-lyase multigene family
in Trifolium subterraneum.";
RL Gene 138:87-92(1994).
CC
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
first reaction in the biosynthesis from L-phenylalanine of a wide
variety of natural products based on the phenylpropane skeleton.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
which is formed autocatalytically by cyclization and dehydration
of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, M91192; AAA17993.1; -
 DR HSSP; P21310; 188F.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGR01226; Phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism; Multigene family.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 211 213 5-imidazolone (Ala-Gly)
 FT MOD_RES 212 212 (By similarity).
 FT FT DHA (2,3-DIDEHYDROALANINE)
 FT (By similarity).
 FT (By similarity).
 SQ SEQUENCE 725 AA; 78998 MW; 70F9925C6622240 CRC64;
 Query Match 28.2%; Score 949; DB 1; Length 725;
 Best Local Similarity 37.2%; Pred. No. 6,7e-66;
 Matches 252; Conservative 108; Mismatches 266; Indels 52; Gaps 18;
 QY 69 LGGVSLTLDVVG-AARKGRXVYVNDSEIRKIDKSVFRLXOLXNSV--YGVTTGFG 125
 DB 68 LGGETLTISQVAIAIAHDGATVEL--SSARAGVASSDWMESMKNKTSYGVTTGFGA 125
 QY 126 SADRTEDAIISQKALLEHQLGVLPSTXDSFPLGRLENSLPLEVYRGANTIRVNSLTR 185
 DB 126 TSHRRKXGQALQKELIRFLNAGIRGNGES-----NHTLPATRAALVIRINLLQ 178
 QY 186 GHSNAPVLVLEALTNFNLNGITPIYPLRGTSASGDLSPITYAATNGHDSKRVNHE 245
 DB 179 GSGIRFELLEIAITKLNNNITPCLPLGTTASGDLVPLSYIGLGRSNSK--AHG 235
 QY 246 GAEKIKXAEALATFGL--EPVVLGPKKEGLVNGTAVASAMATLADHMLSLSQL 303
 DB 236 PGEMLNMAEAQLQAGINAEFFELQPKGIALVNTGAVSGSLAVLFEANILVLEVL 295
 QY 304 TALTVAMTGAHSGFRPLVDVTRPHPTQIEVARNIRFLGSKAFVNHHEEVKVDDEG 363
 DB 296 SAIFAEVWQGR-EPFDHILTKLKHHPGIEAALMEHILHGSAY-VQDAKXLEMDPLQ 353
 QY 364 LRQDDEYPLRTSPQWGLVSDIMIAHVLSEAG-QSTQNPPLDVENKXTHGNGFOA 422
 DB 354 KPRQDRYALRTSPQWGLV--VIRFSTKIEREINSVNDNPLIDVRNALHGSNGFOG 411
 QY 423 SAVANTMEKTRIALALIGKLNFTQTEMNAGNRLPSCL-AAEDPSLYHCKGLDIAA 481
 DB 412 TPVGSMNDTRLALASIGKLPAGSELVNDPYNNGLSNLSASHNPFLDYGFGSEIAM 471
 QY 482 AATTSLEGLHAPVYTHVQPAEMGNQAVNSLALISARRTAENDVLSLLATHLVCLQA 541
 DB 472 ASYCSSELQYLAIPVTHVQSAQHNDVNSLGLISSRKTKEALIEILLQMSSTFLIALCQA 531
 QY 542 VDIRAMEFEKQGFQXXXXLXKHGXXXXTXXXXXXKVKXXLAKRLQGNISYLE 601
 DB 532 IDIRHLEENIKASVKNTVQVAKKTLITIGVSEGLPSRFCELDLKVDR-EHVFSTIDD 590
 QY 602 PRWHDASXATGTVEVEXLSXXXXXVSLAAVNA-----WKVAXAKAAS-LTR 648
 DB 591 P-----CSATYPLAQKLRQ--VIVDHALVNGSEKSNSTISFOKIAFEEELKTLIPK 641
 QY 649 XVPXKXWMAKAPSSSSPALXVLSF-RTRVVXSVFVEELGVYARAGVDFLQKQVITGSNSR 707
 DB 642 EYESAPTAAYENGSTANKINGCRSTPLTKFVEELGT-----SLTGERVISPGEBCDK 696
 QY 708 IVEAKSGRINXVLMQL 725
 DB 697 LFTAMQCGKIIDPLKCL 714

RESULT 11
 ID PAL1_PRUV STANDARD; PRT; 717 AA.
 AC 064963;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase I (EC 4.3.1.5).
 GN PAL1.
 OS Prunus avium (Cherry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
 OX NCBI_TaxID:42229;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Summit;
 RA Wiersma P.A., Wu Z.;
 RT "A full-length cDNA for phenylalanine ammonia-lyase cloned from ripe
 RT Sweet Cherry fruit (Prunus avium).";
 RL (in) Plant Gene Register, PCR98-184.
 CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF036948; AAC78457.1; -
 DR HSSP; P21310; 188F.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGR01226; Phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 DR Lyase; Phenylpropanoid metabolism; Multigene family.
 KW Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 203 205 5-imidazolone (Ala-Gly)
 FT MOD_RES 204 204 (By similarity).
 FT FT DHA (2,3-DIDEHYDROALANINE)
 FT (By similarity).
 FT (By similarity).
 SQ SEQUENCE 717 AA; 77999 MW; B64DF90FA0BF0B3 CRC64;
 Query Match 28.1%; Score 947; DB 1; Length 717;
 Best Local Similarity 35.8%; Pred. No. 9,5e-66;
 Matches 250; Conservative 112; Mismatches 285; Indels 52; Gaps 17;
 QY 40 GSKLPTXTQTQDLYEXKLABPXTDDXKLELDGYS;TLGDVYGARKGVYVXDSEIRX 99
 DB 31 GVAAEFLKGSNHLDEYKRVNAE-YRRPVVRLQGESLTSQVAIAITHDSGVVEVLSGAR 89
 QY 100 KIDKSVFRLXOLXNSV--YGVTTGFGSADRTEDAIISQKALLEHQLCGVLPSTXDSF 157
 DB 90 GYKASDWMDSMNGTDSYGVTTGFGATSHRRTQGAALQKELIRFLNAGVFPSTKES- 148
 QY 158 XLGRGLENSLPLEVYRGANTIRVNSLTRGHSNVR;LVLEALTNFNLNGITPIYVLRGIS 217
 DB 149 -----GHTLPQCATRAALVIRINLLQGYSGIRFEIIEVITKFLNNVTPCLPLRGTT 202

QY 218 ASGDLSTSYIAAATGHPDSKVHYHHEGEXIKMAREALFGLPE--VVLGPEKGL 275
 DB 203 ASGDLVPLSTIAGMLTIRPNSKA-VGPBG--QTLTAABPFEVINGSGFELQPEGLAL 259
 QY 276 VNGTAVASAMATLALDHANMLSLLSQALPTLVEAMVAGAGSFHPLDVTREPHPTOLEV 335
 DB 260 VNGTAVAGSGLASVTLFTDITLITALLSELISAFEVNQGPK-EFTDHLTKLKHGQLEA 318
 QY 336 ARNRTLLGSSKPRVH---HEEEVAVKDEGLRDPRPLTSPQMLGPLVSDMTHAA 391
 DB 319 AALWEHTLDSSSYVKAQKLEEDPLQKP---KQDRALRTSPQMLGPOLE--VIRYS 371
 QY 392 VLSLEAG-QSTNDPLIDVENKXTHGNGFQASAVNTMEKTRMLALIGKLNFTQTEM 450
 DB 372 TKSLERIDSVNDPLIDVSRNKALHGNFGCPPIVSMDNRLTALISGKLMFQPSL 431
 QY 451 LNAQMGFLPSCLA-AEDPSLYHCKGLDIAAAAYTSELGHLANPTVTHVQPAENGQAV 509
 DB 432 VNFYVNGGLPSNLISGGRNPSLDYGFKAELIAMAASYCELQFLANPTVTHVQPAENGQAV 491
 QY 510 NSLALTSARRTAANDVLSLLATLHYCVLQAVDL-PAWFEFKQKQFXXXXLKHGTX 569
 DB 492 NSGLSSKRTAABVDLKMSTFLVALCQALDHLSENLRTNVTNVSQVAKR--T 548
 QY 570 XXTXXXXXXELXXXKXKLKRLQETNSYDLFPRMHDASFATGTVEXLSSXXXXVSL 629
 DB 549 LTTGVNGELHPSRFCEKDLTKVDREYV---AYIDPCSATYVLMQKLQVVEHNT 604
 QY 630 AAVYAKVAKA--EKALSTLTXKXKXFKXAPSSSPALXYSP-----FTRVLYSF 678
 DB 605 NGENKXASISIFQKIVAFEEELKVLPEKVSAAALDGSAGVPRITECRSPYATKF 664
 QY 679 VREELGVKARGDVPFGKQEVTTGSVNSRIYEAIRSGRI 717
 DB 665 VREELG-----AEYLTGKVRSPGECDDKVFALICEGKI 698
 RESULT 12
 PALL PETCR STANDARD; PRT; 716 AA.
 AC P24BI;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
 GN PALL.
 OS Petroselinum crispum (Parsley) (Petroselinum hortense).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Campanulids; Apiales; Apiaceae; Apioidae; apioid superclade;
 OC Apium clade; Petroselinum.
 OX NCBI_TaxID=4043;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89356635; PubMed=2767049;
 RA Lois R., Dietrich A., Hahlbrock K., Schulz M.;
 RT "A phenylalanine ammonia-lyase gene from parsley: structure,
 RT regulation and identification of elicitor and light responsive,
 RT cis-acting elements.";
 RL EMBO J. 8:1641-1648 (1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94326930; PubMed=8050576;
 RA Schuster B., Reley J.;
 RT "Serine-202 is the putative precursor of the active site
 RT dehydroalanine of phenylalanine ammonia lyase. Site-directed
 RT mutagenesis studies on the enzyme from parsley (Petroselinum crispum
 RT L.)";
 RL FEBS Lett. 349:252-254 (1994).

CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (by similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL: X15473; CA33500.1; -
 CC EMBL: X16772; CA34715.1; -
 CC EMBL: Y07654; CA88938.1; -
 CC PIR: S04463; S04463.
 CC DR HSSP: P21310; 188P.
 CC DR InterPro: IPR008948; L-Asparagine-lyase.
 CC DR InterPro: IPR001106; Phe/His NH3lyase.
 CC DR InterPro: IPR005922; Phe_am_lyase.
 CC Pfam: PF00221; PAL; 1.
 CC DR TIGRPFAM: TIGR01226; phe_am_lyase; 1.
 CC DR PROSITE: PS00489; PAL_HISTIDASE; 1.
 CC KM Lyase; Phenylpropanoid metabolism; Multigene family.
 CC FT CROSSLINK 202 204 5-imidazole-one (Ala-Gly)
 CC FT (by similarity).
 CC FT MOD_RES 203 203 DHA (2,3-DIHYDROALANINE)
 CC FT (by similarity).
 CC FT MUTAGEN 203 203 S->A: COMPLETE LOSS OF ACTIVITY.
 CC FT MUTAGEN 210 210 S->A: NO LOSS OF ACTIVITY.
 CC SQ SEQUENCE 716 AA; 77828 MW; 23D82FALAACEFB4A CRC64;
 Query Match 28.1%; Score 946.5; DB 1; Length 716;
 Best Local Similarity 36.7%; Pred. No. 1e-65; Indels 43; Gaps 18;
 Matches 254; Conservative 110; Mismatches 286;
 QY 49 TGLDIVEXXLADPXTDXELDGYSLTGLGVYG-PARKGXVRYXDSERXKIDKSVYE 107
 DB 40 SHLDEVKQWVAE-YRKPVYKLGELTITISQVAAISADGSGVVEISEARAGVXKSPW 98
 QY 108 LKXCLXNSV--YGVTTGGSGADRTEDATSLQKALLHQLGCVLPKSKSFKLGLEN 165
 DB 99 VVDSVKKCTDSYGVTTGGATSHRTKQGGALQKELIRFNAGI-----FNGGSDN 149
 QY 166 SLPLEVVGAMTIRVNSLTRGSAVRLVVLALTNFNLHGITPVPJRGTISAGDLSPL 225
 DB 150 TLPHSATGAAMVAVNLITLQGYSGIRPEILIEATIKFLNQTITCLPARGTITSGDVLPL 209
 QY 226 SYTAATITGHPDSKVHYHHEGEXIKXARALFGLPE--VVLGPEKGLVNGTAVSA 283
 DB 210 SYIAQLTLGRNPSKA-VGPTGV--ILSPFAFKLAGVEGFEPLQPEGLALVNGTAVS 266
 QY 284 SWATLALDHANMLSLLSQALPTLVEAMVAGAGSFHPLDVTREPHPTOLEVARNIRTL 343
 DB 267 GKASWVLFENAILAVLAEVNSAIFAENQPK-EFTDHLTKLKHGQLEAAATVEHIL 325
 QY 344 EGSKFAVHEEEVAVKDEGLRDPRPLTSPQMLGPLVSDMTHAAVLSLEAGSTTD 403
 DB 326 DGSAY-VAAQKLTLEMPPLQKPKDRYALRTSPQMLGPOI-EVIRISTMIEREINSVD 383
 QY 404 NPLIDVENKXTHGNGFQASAVNTMEKTRMLALIGKLNFTQTEMUNAGMNGRGLPSC 463
 DB 384 NPLIDVSNKAIHGNQCPPIGVSMDNRLTALISGKLMFQPSLAVDPTNGGLPSVL 443
 QY 464 A-AEDPSLYHCKGLDIAAAAYTSELGHLANPTVTHVQPAENGQAVNSLALTSARRTAE 522

Db 444 SGRNPISLDYFFKKAELAMASYCELOFLANPVTNNVQSAEQHNQDVNSLGLISSRTSE 503
2y 523 ANDVLSLLATHLVYCVLQAVDLRAMEEFKQFPKXXXLLXQFGKXXXTXXXXLXX 582
Db 504 AVEILKLMSTFFVLGCAIDLRLHEENLKSTVKNVTSVAKXVLTWGVNGELHPSRFE 563
2y 563 KXXKXKLXRELQTSYDLER--RHHDAFSKATCTVVEKXLSXXXXXVSLAAVAMVAXA 640
Db 564 KDLRLVDR-EYTFAYIDDPSCATYPLMOKIRQTLVEHALKNGDNERNLSTSFQKIATP 622
Qy 641 EKAIS--LTRKXKXFWAPSSPALXVLP-----RTRVLYSPREELGVKARGDV 692
Db 623 EDELKALLPKVEESA-RAALESNPAL-----PNRIEGRSPLYKFKRELGT-----EX 672
2y 693 FLGKQVITGSNVRITVPAIKSGRIKXVLMVL 725
Db 673 LTGEKVTSPGEEFKEVFTAMSKGEIIDLPLECL 705
RESULT 13
PAL2_PETCR STANDARD; PRT; 716 AA.
ID PAL2_PETCR
AC P45728;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
PAL2.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
GN Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Campanulids; Apiales; Apiaceae; Apioidae; apioid superclade;
OC Apium clade; Petroselinum.
NCBI_Taxid=4043;
RX MEDLINE=95010141; PubMed=7925471;
RA Appert C., Logemann E., Halbrock K., Schmid J., Amrhein N.;
RT Structural and catalytic properties of the four phenylalanine
ammonia-lyase isoenzymes from parsley (Petroselinum crispum Nym.),";
RU Eur. J. Biochem. 225:491-499 (1994).
CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
first reaction in the biosynthesis from L-phenylalanine of a wide
variety of natural products based on the phenylpropane skeleton.
CC CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
which is formed autocatalytically by cyclization and dehydration
of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: X81158; CAA57056.1; -
DR FRI; S48725; S48725.
DR HSSP; P21310; 188F.
DR InterPro; IPR008948; L-Asparatase-like.
DR InterPro; IPR001106; Phe/His_NH1lyase.
DR InterPro; IPR005922; Phe_His_NH1lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRPFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00486; PAL_HISTIDASE; 1.
KM Lyase; Phenylpropanoid metabolism; Multigene family.
FT CROSSLINK 202 204 5-imidazolone (Ala-Gly)
FT (By similarity).
FT MOD_RES 203 203 DHA (2,3-DIDEHYDROLANINE)

FT (BY SIMILARITY).
SQ SEQUENCE 716 AA; 77964 MW; F74D7BD81BEE66 CRC64;
Query Match 28.1%; Score 946.5; DB 1; Length 716;
Best Local Similarity 35.5%; Pred. No. 1e-65;
Matches 254; Conservative 110; Mismatches 264; Indels 87; Gaps 19;
Qy 49 TQDLIVEXXLADPXTDDXELDGYSLTLDGVG--AARKGRVXVXDSDERXKIDKSVR 107
Db 40 SHLDEVKQMAVE-YRKVVVLGGETLITISQVAAISARDGSGVTVLSEARAGVAKSSDW 98
Qy 108 LRQOLXNSV--YGVTTGPGGSADTRTEDAISLOKALLEHQLCGVLPFSXDSFYLGGLEN 165
Db 99 VMDSMNKGTDSYGVTTFGATSHRRTRQOGALQKELIRFLNAGI-----FGGSDN 149
Qy 166 STPLEVYRGAMTIVNSLITRGHSAVRVLVEALTNFNGITIVPLRGITISASGLSP 225
Db 150 TLFHSATRAAMLVINILLOGYSIGIRFEIIEATKPLNQTICLPRLGITTSGLDVL 209
Qy 226 SYIAAATGHDPSKXVYHGEKXIMKAREALALFGLER--VVLGREGIGLVNGTAVSA 283
Db 210 SYIAGLITGRNSKA-VGPTGV--ILSPEEAFKLAVGEGFPELOPREGIALVNGTAVGS 266
Qy 284 SMATIALHDAMLSLLQALTLALVEMVGHAGSFHELDVTRPPTQTEVARNITL 343
Db 267 GVASVTFEANTILVLEWMSAIPAEVMQKP-EFTDHLTKLKHNGQTEALAIMEHIL 325
Qy 344 EGSKPFAVHHEEEVVKRDEGILRDRYPLRTSPQMLGVLSDMIIHAHVLSLAGOSTTD 403
Db 326 DOSAY-VKAAKLMHMDPLQKPKQDRVALRTSPQMLPQ--EVIKRSSTKMIEREINSVND 383
Qy 404 NPLIDVENKXTHHGGNFGASAVXNTMEKTRALALTKLNFQTLTEMLNQMNRGJPSCL 463
Db 384 NPLIDVSRNKAIHGGNFGCTPIGSMNDTRALAIATKLMFAQSELVNDPYNGLPSNL 443
Qy 464 A-AEDPSLSYHCKGDIILAAAYTSELGLANPVTTHYOPAEEMGNQAVNSLALISARETAE 522
Db 444 SGRNPISLDYFFKKAELAMASYCELOFLANPVTNNVQSAEQHNQDVNSLGLISSRTSE 503
Qy 523 ANDVLSLLATHLVYCVLQAVDLRAMEEFKQFPKXXXLLXQHF-----GXXTXXXX 577
Db 504 AVEILKLMSTFFVLGCAIDLRLHEENLKSTVKNVTSVAKXVLTWGVNGELHPSRFE 563
Qy 578 KKLXKXKXKX-----LKKREQT-----NSYDLEPRHDAFSXATGVV 616
Db 564 KDLRLVDR-EYTFAYIDDPSCATYPLMOKIRQTLVEHALKNGDNERNLSTSFQKIATP 623
Qy 617 EXLSSXXXXXVSLAAVAMVAXAKAISLTRVXKXFWAPSSSSPALXVLP----- 670
Db 624 DELK-----ALPKVESAR-----AALBSGNPAL-----PNRIEBC 655
Qy 671 RTRVLYSPREELGVKARGDVFLGKQVITGSNVRITVPAIKSGRIKXVLMVL 725
Db 656 RSYPLKFKRELGT-----EYLTGEKVTSPGEEFKEVFTAMSKGEIIDLPLECL 705
RESULT 14
PAL1_SOYBN STANDARD; PRT; 713 AA.
ID PAL1_SOYBN
AC P27991;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Phenylalanine ammonia-lyase 1 (EC 4.3.1.5).
PAL1.
OS Glycine max (Soybean).
GN Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OC NCBI_Taxid=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. T225; TISSUE=leaf;

EX MEDLINE=92190550; PubMed=1799682;
 RA Frank R.L., Yodkin L.O.;
 RT "Sequence and structure of a phenylalanine ammonia-lyase gene from
 RL DNA Seq. 1:335-346(1991).
 CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide
 CC variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine + trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PIM: Contains an active site 4-methylcinnamylidene-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (by similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL; X52953; CAA37129.1; -.
 DR PIR; S22991; S22991.
 DR HSP; P21310; 1B8F.
 DR InterPro; IPR008948; L-Aspartase-like.
 DR InterPro; IPR001106; Phe/His_NH1Lyase.
 DR InterPro; IPR005922; Phe_am_Lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KM Lyase; Phenylpropanoid metabolism; Multigene family.
 FT CROSSLINK 199 201 5-imidazolinone (Ala-Gly)
 FT MOD_RES 200 200 (By similarity).
 FT MOD_RES 200 200 DHA (2,3-DIHYDROALANINE)
 FT (By similarity).
 SQ SEQUENCE 713 AA; 77744 MW; 9D71EF1CC230216A CRC64;
 Query Match 28.1%; Score 946; DB 1; Length 713;
 Best Local Similarity 36.6%; Pred. NO. 1,1e-65;
 Matches 271; Conservative 105; Mismatches 293; Indels 72; Gaps 23;
 QY 14 NGXXNGXHAAXXASXXXXXXAXAGSXLPTTXXTQDIVEXXADLPXTDXXELDYS 73
 DB 5 NGHONGSFCGLSTAKGNNDPLNWCANA---EAMKSHLDEVKENVAE-YRKPVRLGGERT 59
 QY 74 LITGDVVAARKGRKYRVXDSEITXKIDKSYFIRKXOLXSV--YGTTFGFGSADTRT 111
 DB 60 LITQAQAAVAGHDGVAELSESAREGVKASSEWMSNMNGTDSYGTTFGFGATSHRR 119
 QY 132 EDAISLOKALIEHOLCGVLPSTXDSFXLGRGLENS--LPLEVGAMTIRVNSITRGSA 189
 DB 120 KCGGALQKELIRFINAGI-----FNGTGESHTLPHRATMAALVAVINTLLQYSG 170
 QY 190 VALVLEALTNLNGITPIYVLRGTIASGDLSPSTIAAATGHPDSKVYHXHEGXEK 249
 DB 171 IFEILEAITKLNNNVNPPCDLROGTITASGDLVPSYIAGLITGRNSXA-VQPSG-E 227
 QY 250 IXXAREBALFGL--EPVVLGPKEGGLVNGTAAVSASAMTLADHMLSLTSQLTALT 307
 DB 228 VINAKAEPELASINSEFELQPKBELALVNGTAVSGSLASVLPFANILVLESEVSAIF 287
 QY 308 VEAATGAGSGFHPFHDVTRPHPTQIEVARNIRFLGSGFAYV---HEEEVKVXDDEG 363
 DB 288 AEVWGKRP-EFTDHLTHRLKHPQIEEAALMEIILDGSSYMRKAKKHLIDLPQRP-- 343
 QY 364 ILROORYLRTSPQGLVSDMTAAHAAVLSLENG-QSTTDNPLIDVENKXTHGNGFOA 422
 DB 344 --KORYALRTSPQWLPLE--VIRSTSTIESEINSVDNPLIDVSRKALHGNFGG 399
 QY 423 SAVXNTEKRLALALIGKLNFTOLTETMLNAGMNRGDPGL-AAEDPSLSYHCKGLDIAA 481

RESULT 15
 ID PAL4_ARATH STANDARD; PRT; 707 AA.
 AC Q9SS45;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable phenylalanine ammonia-lyase (EC 4.3.1.5).
 GN AT3G10340 OR F14P13.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unsel D.,
 RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA De Simone V., Boutry M., Grivell L.A., Meche R., Pigdonmech P.,
 RA Winkler P., Cattolico L., Artiguenave F., Robert C., Brotier P.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Dizonet H., Erile H., Jordan N., Bangert S.,
 RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
 RA Reichelt J., Schaefer M., Schoen O., Bagues M., Terol J., Clement J.,
 RA Navarro P., Collado C., Perez-Perez A., Oetemaelder B., Duchemin D.,
 RA Cooke R., Laude M., Berger-Llauro C., Puntelle B., Maury D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casasuberta E.,
 RA Monfort A., Argitjov A., Flores M., Liquri R., Vitale D.,
 RA Monhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., News H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walte A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Miltischer J., Sellers P., Gill U.E., Feldblum T.V.,
 RA Preuss D., Lin X., Niemann W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Ideasa K., Kawashima K., Kishida Y.,
 RA Kiyokawa S., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shindo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RA "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 RT thaliana."
 RL Nature 406:820-822(2000).
 CC -1- FUNCTION: This is a key enzyme of plant metabolism catalyzing the
 CC first reaction in the biosynthesis from L-phenylalanine of a wide

CC variety of natural products based on the phenylpropane skeleton.
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC -1- PATHWAY: Phenylpropanoid biosynthesis; first step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residue Ala-Ser-Gly (By similarity).
 CC -1- SIMILARITY: Belongs to the PAL / histidase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AC009400; AAP02809.1; -.
 CC HSSP; P21310; 1B8F.
 CC InterPro; IPR008948; L-Asparatase-like.
 CC InterPro; IPR001106; Phe/His NH3lyase.
 CC InterPro; IPR005922; Phe_am_lyase.
 CC Pfam; PF00221; Pal; 1.
 CC TIGRfam; TIGR01226; phe_am_lyase; 1.
 CC PROSITE; PS00488; PAL_HISTIDASE; 1.
 CC Lyase; Phenylpropanoid metabolism; Multigene family.
 CC CROSSLINK 194 196 5-imidazolone (Ala-Gly)
 CC (By similarity).
 CC MOD_RES 195 195 DHA (2,3-DIDEHYDROALANINE)
 CC (BY SIMILARITY).
 CC FT SEQUENCE 707 AA; 76919 MW; B67743C941B5089B CRC64;

Db 567 FSYADDPCLRYPLMOKLRIHLVDHALADPEREANSATSVFHKIGAFE-----AEIKLL 620
 Qy 646 LTRXY-RXXFMXAPSSSPALXYISPTTRVLYSFVRELGVKARGDPFLCKQEVITGSN 704
 Db 621 LPKEVERVRVEYEGSTAIARIRICECSYPLRYFRQDELNT-----ELLGENVRSQEE 675
 Qy 705 VSRIVEAIKSGRINXVLVWML 725
 Db 676 FDKVFLAISDGKLDPLPLECL 696
 Search completed: September 9, 2004, 10:04:42
 Job time : 19.5685 secs

Query Match 28.0%; Score 943.5; DB 1; Length 707;
 Best Local Similarity 37.0%; Pred. No. 1.7e-65;
 Matches 252; Conservative 98; Mismatches 282; Indels 49; Gaps 16;
 Qy 65 DXXELDGYSLTLDGVGAARKGRVYKXSDIRKIDKSVFLRXQ.XNSV--YGVTTG 122
 Db 45 EAVKLGGETLTIGQVAAVARGGGSTVELAEARAGVKAASSEWMSNRGTDSTGYVTTG 104
 Qy 123 FGGSADPTREDALISLOKALLEHQLCGVLPTSDSFXLGRG--LENSLPLEVVRGAMTIRVN 181
 Db 105 FGATSHRRTKOGALONELIRFLNAGI-----FGPGAGDTSHTLPKPTRAAMLVNVN 157
 Qy 182 SLTRGSAVRLVLEALTNFLNAGITPIVPLRGTTASGDLSPSLYIAAATGHPDSKVH 241
 Db 158 TLLQGYSGIRFETILEAITKLINHEITPCPLRGTTASGDLVPLSYIAGLLTGRPNKA- 216
 Qy 242 VMEGEEKIMXAREATLALFGLPEVY--LGPKEGLGVNGTAVASMATLALHDAHWLSTLS 300
 Db 217 VGRSG--ETITASEAFKLAGVSSFFELQPKREGIALVNGTAVGSLASTVLPDANILAVLS 274
 Qy 301 QALTALTEAMVGHAGSFHFPLHDVTRPHPTQIEVARNIRTLLEGSKFA---VHHEEV 356
 Db 275 EVMSAMFAEVMGKR--EFTDHLTHKLKHHPQIEAAAIIMEHILDGSSYVKEAQLHEMDP 333
 Qy 357 KYVDDDEGLKQDEYPIRTSPQWLGPLVSDMIHAHVLSLBAGQSTNDPLIDVENKXTHH 416
 Db 334 LQRP-----KQDRVALRTSPQWLGQPI--EVIRATKMTIREINSVNDPLIDVSRNKALH 387
 Qy 417 GGNFOASAVXNTMEKTRIALALIGKLNFTQLEMLNAGNNGLPSCLA--ABDPSSLSTHCK 475
 Db 388 GGNFOGTPPIGVANDNSRLALASIGKLMFAQFSELVNDPYNGLPSNLSGGRNPSLDYGF 447
 Qy 476 GLDIAAAAYTSELGHLANPTTHVQPAEKGNAVNSLALISARTEANDVLSLLATHL 535
 Db 448 GARIAMASYCESELQFLANPVTHVQSAEQHNDVNSLGLISSRKTAEADVILKLMSTYL 507
 Qy 536 YCVLQAAVDLRAEMEFPKQFPXKXXLLKQHFQXXXKXXKXXKXXKXLEQT 595
 Db 508 VALQAAVDLHRLHENLKKAAVKSASQAAKAVLTVGANGELHPSRFTERYDLYQVDR--EYV 566
 Qy 596 NSYDLEP-----RWDAPSKATGVTVEXLSSKXXVSLAAVNAKVAAXEKAIS 645

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2004, 09:57:28 ; Search time 88.3651 Seconds
(without alignment)
2592.269 Million cell updates/sec

Title: US-09-939-408a-21
Perfect score: 3365
Sequence: 1 NAPSLSLSTATXANXNGX.....RIYAIKSGRLNVLVLM 726

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP_archaea:*
- 2: SP_bacteria:*
- 3: SP_fungi:*
- 4: SP_human:*
- 5: SP_invertebrate:*
- 6: SP_mammal:*
- 7: SP_mhc:*
- 8: SP_organelle:*
- 9: SP_phase:*
- 10: SP_plant:*
- 11: SP_protist:*
- 12: SP_virus:*
- 13: SP_vertebrate:*
- 14: SP_unclassified:*
- 15: SP_virus:*
- 16: SP_bacteriophage:*
- 17: SP_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	969	28.8	725	10 Q94ANI	Q94ANI arabidopsis
2	967	28.7	711	10 Q8W2B4	Q8W2B4 lactuca sat
3	965.5	28.7	707	10 Q7X7G3	Q7X7G3 oryza sativ
4	963	28.6	671	10 Q84VRO	Q84VRO oryza sativ
5	958.5	28.5	703	10 Q8VXG7	Q8VXG7 zea mays (m
6	955	28.4	711	10 Q9AX15	Q9AX15 paribitis n
7	955	28.4	730	10 Q9W567	Q9W567 rubus idaeu
8	954	28.4	708	10 Q94ENO	Q94ENO rehmannia g
9	944.5	28.1	681	10 Q8RUZ3	Q8RUZ3 pinus sylve
10	944.5	28.1	715	10 Q852S1	Q852S1 daucus caro
11	943.5	28.0	681	10 Q8SAS6	Q8SAS6 pinus sylve
12	942.5	28.0	716	10 Q8WAX1	Q8WAX1 catharantu
13	941.5	28.0	681	10 Q8RW49	Q8RW49 pinus sylve
14	940	27.9	717	10 Q8RWP4	Q8RWP4 arabidopsis
15	939	27.9	713	10 Q8GZR8	Q8GZR8 lactuca sat
16	937.5	27.9	681	10 Q8SAS7	Q8SAS7 pinus sylve

17	931	27.7	710	10 Q94C45	Q94C45 manihot esc
18	929.5	27.6	716	10 Q94IN7	Q94IN7 agastache r
19	926.5	27.5	721	10 Q9XFX5	Q9XFX5 citrus clem
20	922	27.4	687	10 Q94F90	Q94F90 manihot esc
21	919.5	27.3	727	10 Q7XJC3	Q7XJC3 pinus pinas
22	917.5	27.3	727	10 Q7XJC4	Q7XJC4 pinus pinas
23	915.5	27.2	714	10 Q7XCM8	Q7XCM8 oryza sativ
24	914	27.2	711	10 Q8H6V5	Q8H6V5 populus tre
25	911.5	27.1	714	10 Q8H6V6	Q8H6V6 populus tre
26	910	27.0	717	10 Q8H6W0	Q8H6W0 coffea cane
27	904.5	26.9	712	10 Q94F89	Q94F89 manihot esc
28	900	26.7	655	10 Q9XG83	Q9XG83 vigna ungu
29	898.5	26.7	549	10 Q04876	Q04876 hordeum vul
30	895.5	26.6	718	10 Q9XFX6	Q9XFX6 citrus clem
31	893	26.5	703	10 Q841I2	Q841I2 phalaenopsi
32	893	26.5	715	10 Q24286	Q24286 populus kit
33	892.5	26.5	619	10 Q39545	Q39545 cucumis mel
34	887	26.4	619	10 Q8H6V9	Q8H6V9 coffea cane
35	868.5	25.8	618	10 Q9FS00	Q9FS00 dianthus ca
36	763	22.7	396	10 Q8WVP6	Q8WVP6 beta vulgar
37	736.5	21.9	544	10 Q9FT38	Q9FT38 betula verr
38	735	21.8	532	10 Q81647	Q81647 capicum ch
39	717	21.3	380	10 Q84U27	Q84U27 ginkgo bilo
40	715	21.2	362	10 Q9FU17	Q9FU17 rubus idaeu
41	703	20.9	363	10 Q9FU18	Q9FU18 rubus idaeu
42	612.5	18.2	497	10 Q04869	Q04869 hordeum vul
43	604.5	18.0	475	10 Q04875	Q04875 hordeum vul
44	600.5	17.8	436	10 Q9LEP1	Q9LEP1 betula verr
45	596.5	17.7	435	10 Q9M3U1	Q9M3U1 betula verr

ALIGNMENTS

RESULT 1
Q94ANI PRELIMINARY; PRT; 725 AA.

AC Q94ANI;
ID Q94ANI;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia lyase (Hypothetical protein) (EC 4.3.1.5).
GN At2G37040 OR PAL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids; Il; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_Taxid=3702;

RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Yu S., Bower L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Kosena E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shim P., Southwick A., Tracy S.E., Shinozaki K., Davits R.W.,
RA Ecker J.R., Theologis A.;
RT "Full length cDNA of gene TlJ8.22/At2G37040 (GI:6598547).";
RN Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Nguyen M., Tripp M., Palm C.J., Jones T., Wu T.,
RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,
RA Dang J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,
RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Shim P., Tang C.C., Toriumi M., Wong C., Wu H.C.,
RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,
RA Davis R.W.,
RN Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etku P., Lee J.M.,

RA Tortolini M., Yu G., Brooks S., Chao Q., Chen H., Kerlin-Neumann G.,
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
RA Southwick A., Davis R.W., Ecker U.R., Theologis A.,
RL "Arabidopsis Open Reading Frame (ORF) Clones",
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cochrane F.C., Kim M.R., Cardenas C.L., Davin L.B., Lewis N.G.;
RT "Putative PAL multigene family in Arabidopsis thaliana: Kinetic
RT Characterization of Recombinant Proteins and Knock-out Mutant
RT Characterization";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY045919; AAK76593.1; -
DR EMBL; BT003330; AAC02949.1; -
DR EMBL; AY079363; AAL85094.1; -
DR EMBL; AY031128; AAP59436.1; -
DR PIR; G84787; G84787.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:00162.1; Pantoic acid lyase activity; IEA.
DR GO; GO:0016841; Pantoic acid lyase activity; IEA.
DR GO; GO:0016829; Pantoic acid lyase activity; IEA.
DR GO; GO:0009058; Pantoic acid lyase activity; IEA.
DR GO; GO:0006559; Pantoic acid lyase activity; IEA.
DR InterPro; IPR008948; Pantoic acid lyase; IEA.
DR InterPro; IPR001106; Pantoic acid lyase; IEA.
DR InterPro; IPR005922; Pantoic acid lyase; IEA.
DR Pfam; PF00221; PAL; 1.
DR TIGRfam; TIGR01226; Pantoic acid lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KM Hypothetical protein; Lyase.
SQ SEQUENCE 725 AA; 78725 MW; 0262653B2DEF9CE CRC64;

Query Match	28.8%;	Score 969;	DB 10;	Length 725;
Best Local Similarity	35.5%;	Pred. No. 8.8e-68;		
Matches 258;	Conservative 120;	Mismatches 257;	Indels 92;	Gaps 21

```

QY      40 GSKLPFTTXXOLDIVEXXLADPXTDDXHELOSYSITLGDVVGARKRXYRYMDSBETRX 99
Db      39 GAAAEQMKSGSHDEVKRYAE-FRKPVNNLGGSTLTTCQVAISTGNSVAVELSEFARA 97
QY      100 KIDKSYELRXQLXNSV-YGVTTGFGSADTRIEDAISIOLKALLEHQLOGVLPSTXSF 157
Db      98 GVAASDDWVESMNKCKTDSYGVTTGFGCATSHRRTNQVALCKELIRFEMAGIFGSTRET- 156
QY      158 XLGRGLSNSLPLEVVAGMTIRVNSLTGHSVAVRIVVEALTNELNGITPIVPLRGTIS 217
Db      157 -----SHLPFSAIRPAAMLVRIINTLLQFSGIREIIEAITSLENNNTTSPSLRKTIT 210
QY      218 ASGDLSPLSTYIAAATGHPDSSKVVHVXHBQEXEKIMXAREALFGLER-VTLGPKBGL 275
Db      211 ASGDLPPLBYIAGLLTGRENPKA-TGPNQ--EALTAEEAFKLAGISGFDLPQKEGIAL 267
QY      276 VNSTAVASAMATLAHDAMLSLSQATLTALVEAMVHAQSFHPFDVTRPHPIQIEV 335
Db      268 VNGTAVGSGMAQVLPETVIVSVLAEILSAVPAEMSKP-EFDHDLTHLKHHPQIEA 326
QY      336 ARNIRTLFGSKF---AVHHEEVKXVDGEGILRQDRPYLRTSPQWLGPIVSDMITHANA 391
Db      327 AALMEHILDGSSYMKLAQKLEHMDLPQK-----KODRYALRTSPQWLGPOIE--VLRPA 379
QY      392 VLSLEHG-OSTDNPLDIVENCKTHHGNGFOASAVXNMEKTRIALILKGNFTOLTCK 450
Db      380 TKSIEHEINSVANDNPLDIVSRKKAHGNFQGTPIGVSMQDTRIALIAIKLMPAQFSEL 439
QY      451 LNAQNRGLPSC-LAAEDBSLYHCKGLDIAAAVTSLEHLANPTVTHVQAPEMGNQAV 509
Db      440 VANDFYNNQPPSNLTASRNPSELVYGFKGAEIMASYSCEQLANPVSHTVQSAEQRHODY 499
QY      510 NSLTALISARTAEADVLSLLATLHYCYLDAVDIRAMEFEFFKQFPXXXXLXHQFGX 569
Db      500 NSLGLISSKRTSEADVILKLTMETTLVAICQAVDRHLEURLQ-----543
QY      570 XTXXXXXXELXXKY-----KKYXKLRL--EQTNSYLEFRMTADPASKAT 612

```

```

Db      544  --IVKNTVSVAKGVLTITGVNGSELHPSRCECDLKKVVDREQVYIYADP-----CSAT 598
OY      613  GTVEVEXSSXXXXXVSLAANA-----MFAVAEKAIS--LTRXVRXXFWXAPS 655
Db      596  YPIPIOKLRQ---YIVDHALINSESEKNVATSIFFHIGAFEEELKAVFKEVEAARAAYDN 652
OY      660  SSSPAPLXYLSP--RTVLVSGFVEEELGVKARQDGVFLGQEVATIGSNVRIYEAIKSRIN 718
Db      653  GTSAIPNRIIECHSGLYLVFVREJGT-----ELLTGKVTSPGEEFPKVTALCEKII 707
OY      719  XVLVKRL 725
Db      708  DPMMECL 714

```

RESULT 2
Q8W2E4
ID Q8W2E4 PRELIMINARY; PRT; 711 AA.

DT 01-MAR-2002 (TREMBlre). 20, (Created)
DT 01-MAR-2002 (TREMBlre). 20, (last sequence update)
DT 01-OCT-2003 (TREMBlre). 25, (last annotation update)
DE Phenylalanine ammonia-lyase.
OS *Lactuca sativa* (Garden lettuce).
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; asterids
OC Campanulids; Asterales; Asterales; Asteraceae; Cichorioideae; Cichorieae;
OC Campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae;

0C LACTUGA,
0X NCBI_TaxID=4236;

RN [1]
RN
SEQUENCE FROM N.A.
RN
Comments: "Washed" by Carl J. E. Colburn's
D

RT "Characterization of phenylalanine ammonia-lyase (PAL) gene in wounded tobacco leaf tissue"

RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF296870, AF296871, AF296872, AF296873

DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0016211; P:plasma membrane; IEA.
DR EMBL AT255350; A:ALU5242.1; -.

DR GO:0016841; F:ammonia-lyase activity; IEA.
DR GO:0016829; F:lyase activity; IEA.

DR GO:0006559: P:phenylalanine catabolism; TRA
DR GO:0009058; P:biosynthesis; IEA
DR GO:0000027; A:gene expression; IEA

DR InterPro: IPR008948; L-Aspartase-Like.
DR InterPro: IPR001106; Phe/His NH3iyase.
DR

DR InterPro: IPR005922; Phe_am_lyase.
DR Pfam: PF00221; PAL; 1.

DR TIGRFAMS; TIGR01226; phe am lyase; 1.
DR PROSITE; PS00488; PAL HI $\overline{\text{ST}}$ IDIASE; 1.

KM Lyase. -
 SQ SEQUENCE 711 AA: 77363 MW: 3120AB2877C29513 CRC64:

Query Match 28.7%; Score 367; DB 10; Length 711;

Best Local Similarity 36.9%; Pred. No. 1.2e-67;
Matches 263; Conservative 108; Mismatches 262;
Indels 80; Gaps

49 TQLDIVEXLADPTDDXXELDGYSLTLGDVG--AARKGRXRVXDSDEIRXKIDKSYE

Dd 32 SHLDEVKKNVAE-FRKRPVYKLGGETLTVSQVAGTAAANDSDITYKVELSEARAGVKASD

107 FLRXQLNSV--YGVTTGFGSADRTTEDAISIQKALLEHQLCGVLP TSDSFXLGRGHE

Ddb
91 WVNESMNKGTDSYGVTFGATSHRRITQGGALQKELIRFLNAGI-----FGNGTE

165 NS--LPLEVRGAMTIRVNSLTGRGSAVRLVLEALTFLNHGITPTVPLRGTISASGDL

Db 142 TSHTEHSATRAAMIVRINTLLQGSIGRFEILAEITKELNNNTTPCPLPRTITASGL

223 SPISYIAAATGHPDSKTVHXHEGEXKIMKARAIALFGLP--VVLGPREGGLVNGTA

Db 202 VPLSYIAGLLTGRPNska-VGPTG--EVLNAEKFAAAGVEGGFELQPKEGIALVNGTA

281 VSASMTLALHDAHMLLSQALTALTVEAMVGHAGSFHPLHDVTRPHTQIEVARNR

```

Db      259 VSGSNASWFLPFAHVALALISEVLSIFAEVMOGRK--ETDYLTKLKHHPQIEAAIME 317
QY      341 TLLEGXFVAVHHEEVKKVXDDEGILRQDRYPLRTSPQWLGPLVSDMTIAHAAVLSLEAGQS 400
Db      318 YLIDGSDY--VKAQKQKHEMDPLQKPKDRVALRTSPQWLGPOI--EVIRSTRKMIEREINS 375
QY      401 ITDNPILDVENKXTHHGNGFQASAVXNMTEKTRIALALIGKLNFTQLTETMNAAGNGLP 460
Db      376 VNDNPILDVSRNKAHLHGNGFQGTPIGVSDNTRILALAIKLMFAQISELVNDPYNNGLP 435
QY      461 SCIA--AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISAR 519
Db      436 SNLSGGRNPSLDYGFGRGELIAMAASCESELQFLANPVTNHVQSAEQHNDVNSLILISARK 495
QY      520 TAEANDVLSLLATHLYCVLQAVDLRAMEFEFKQFXXXXLXQHF-----GXHTXX 574
Db      496 TAEAVDILKMSSTYLVALQSIDLRHLENNKSTVXNVTQVAKVLTGVSNEELHPSR 555
QY      575 XXXXELXXVXXX-----LXKRLQGT-----NSYDLPRWHDAPFSXATG 613
Db      556 FCKKDLIRVDEHYFATIDVCSGTPLMOKLRQVLVDHALNGETEKNNTSIFQKIA 615
QY      614 TVEXLSXXXXXVSLAAVANAKVAAAEKALSITKXVXXFWXAPSSSPALXLSPRTR 673
Db      616 TFEELK-----VLLPKEVEGVRIAYENDTLISIPRIK-----ACRSY 653
QY      674 VLYSFVFEELGVKARGDVEFLKQEVTL--GSNVSYIEAKISGRINXVLYKVL 725
Db      654 PLIRFVRELG-----RG--FLTGEKYSFGEEFDFVFTAMCKGQIIDPLECL 700

RESULT 3
ID Q7X7G3 PRELIMINARY; PRT: 707 AA.
AC Q7X7G3;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DE OSJNB0073E02.12 protein (OSJNB0061C13.1 protein).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu Z., Chen L., Fan D.L., Meng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.D., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Gu W.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.F., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RA Submitted (WAY-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL731616; CA05452.1; -
DR EMBL; AL731629; CA05619.1; -
DR SEQUENCE 707 AA; 7603 MW; 841E4C6D077B1AA CRC64;

Query Match 28.7%; Score 965.5; DB 10; Length 707;
Best Local Similarity 37.1%; Pred. No. 1.6e-67;
Matches 257; Conservative 122; Mismatches 273; Indels 41; Gaps 20;

49 TOLDIVEXXADXTDXKELDGYSLTLGPV--VGAARGRVYRVXSDPEIKXKIDKVE 106
Db      29 SHDDEVKRWAAQ--SRENVKITESSLSLRVGVAVAAKASGVVVELDEARPRKASE 87
QY      107 FLRXQLXN--SVYGVTTGGGSAADRTEDAISIQKALLHQJCGVLPSTXSDFXLGRGLE 164
Db      88 WILNCIALHGDIYGVTTGGTSHRRTKQGAOLQVLELRLNLNAGIFGNGSDG----- 139

```

```

QY      165 NSLPLEVYRGAMTIRVNSLTGRGSAVRLVLEALTNTFNHNGITPIVPLRGTSISAGDLP 224
Db      140 NSLPSEVRAALVAINTLLQIGIRFEILATKILNVSVCLELRGITTSGLVP 129
QY      225 LSYIAAATGHPDSKVYHHEGKEXKAKAREALALFGLP--VVLGREGSLGVNGTAVS 282
Db      200 LSYIGLITGRPNQA--VTVDG--KKVDAEAFAKIAGGFFRLEBEKEGIALVNGTSVG 256
QY      283 ASMATLALHDAMLSLSQALTALTVEAMVGHASFFPLHDVTREPTQTEVARNRTL 342
Db      257 SALAMVLYDANVALVSEVLSAVFCEVNGKPEYTHLTKL--KHHPGSTEAAAIHEHI 315
QY      343 LEGSXFVAVHHEEVKKDDEGIL--RQDRYPLRTSPQWLGPLVSDMTIAHAAVLSLEAG-Q 399
Db      316 LAGSAPFWHAQ--KNEVVDPLKPKQDRVALRTSPQWLGPOI--EVIRA--ATKSIEREVN 370
QY      400 STTDNPILDVENKXTHHGNGFQASAVXNMTEKTRIALALIGKLNFTQLTETMNAAGNRL 459
Db      371 SYNDNPVIDVHRGKALHGNGFQGTPIGVSDNTRILALAIKLMFAQFSELVNEFYNNGL 430
QY      460 PSCIA--AEDPSLSYHCKGLDIAAAAYTSELGHLANPVTTHVQPAEMGNQAVNSIALISAR 518
Db      431 TSNLGSGRNPSLDYGFGRGELIAMAASCESELQFLANPVTNHVQSAEQHNDVNSLIGVSAR 490
QY      519 TAEANDVLSLLATHLYCVLQAVDLRAMEFEFKQFXXXXLXQHFXXXXXX 578
Db      491 TAEAVDILKMSSTYLVALQAVDLRHLLENNKSAVNCVTTAKKVLTTGPAGLHSA 550
QY      579 ELXXVYXKXKRLQGTNSYDLR--RWHDAPFSXATGVEXLSXXXXXVSLAAVANAK 636
Db      551 RPSERKALTAIDR--EAVYSYADDPSCANYPLMTKIRAVLVBALANGAECDSSVPSK 609
QY      637 VAAAEKAI--SLTRVXRXFWXAPSSSPALXYL--SPRTVLYSFVRELGVKARGDVF 693
Db      610 ITAFEEELREALPREMEARVAFETGTAPITNRKESRSPLYRFVRELGC-----VY 663
QY      694 L-GKQEVTTGSAVSRITAEIKSIRINXVLYKVL 725
Db      664 LTGEKLSPEGBECNKFALISERKLIIDPLECL 696

RESULT 4
ID Q84VE0 PRELIMINARY; PRT: 671 AA.
AC Q84VE0;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase.
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipomdare;
RA Cooper B., Hutcheson D., Park S., Guimil S., Lugimbuhl P., Eliero C.,
RA Goff S., Glazebrook J.;
RA "Identification of Rice (Oryza sativa) Proteins Linked to the Cyclin-
RT Mediated Regulation of the Cell Cycle."
RT Submitted (JAN-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY224546; AA072666.1; -
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; Fatty acid ligase activity; IEA.
DR GO; GO:0016841; Fatty acid ligase activity; IEA.
DR GO; GO:0016829; Fatty acid ligase activity; IEA.
DR GO; GO:0009058; P-biosynthesis; IEA.
DR GO; GO:0006559; P-biosynthesis; IEA.
DR InterPro; IPR008946; L-Asparagine catabolism; IEA.
DR InterPro; IPR001106; Phe/His NH3 lyase.
DR InterPro; IPR005922; Phe_am_lyase.

```


DR Pfam: PF00221; PAL; 1.
 DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 671 AA; 72277 MW; 23F93B995CE21E60 CRC64;

Query Match 28.6%; Score 963; DB 10; Length 671;
 Best Local Similarity 37.5%; Pred. No. 2.4e-67;
 Matches 253; Conservative 118; Mismatches 263; Indels 40; Gaps 19;

QY 68 ELDSYSLTLDV--VGAARKGRVYVXSDELPRKIDKVEFLPKOLN--SVYGVTTGF 123
 DB 11 KIEGSSLRVGVAAVSAAXDASGVVLEDEAPRKASSEMILNCIANGDIYVTTIGF 70
 QY 124 GGSADRTEDALISQKLLHQLCVLPFSXDSFKLGRLENSLPLEVVRGAMTRVNSL 183
 DB 71 GGTHRRYKQGAQLQVELLRHLAGLFGNGSDG-----NSLSEVSRAMLVRIINTL 122
 QY 184 TRGSAVALVLEALTNFLNGITPIVPLRGTTISAGDLSPLSYAAATGHPDSKVAVX 243
 DB 123 LQGSYGRFELILEITLINTGVSPLRGTTISAGDLSPLSYAAGLITGRPNAQ-VT 181
 QY 244 HEKEXKIMKREPAIALFGLEP--VLGPYEGGLVNGTRAVASAKMTALIDHAMLSLSQ 301
 DB 182 VDG--KKVDAAEAFKTAGIQGGFFRLPEKGLAIVGTSVGSALAMVLYDANVLAJSE 239
 QY 302 ALTALTVEMVGAAGSFHPLHDVTRPHPTQIEVANRITLLEGSXPAVHHEEVKVKD 361
 DB 240 VLSAVPCEVWNGKREYTDHLLTKL-KHFGSIEAALIMHILAGSAFPHQO--KQNEV 295
 QY 362 EGIIL--RQDEYPLRTSPQWLGPLVSDIMAHAVLSLEAG-QSTTDNPLIDVENKXTHHG 418
 DB 296 DPLPKQDXYALRTSPQWLGPGI-EVIRA-ATKSIEREVSVDNPFVIDVHRGKALHG 353
 QY 419 NPGASVXNMTKMTLALIGLNTQUTLTMANMGRGLPSCLA-AEDPSLSYHCKGL 477
 DB 354 NFGQTPGVSMQDTRALIANIGLMAFQSELVNFYNGILSNLAGSRNSLDYGFPGT 413
 QY 478 DIAAAVTSLEGLANPVTHVQPAEMGNQAVNSLALISARRTAEANVLSLLATHLYC 537
 DB 414 EIMASYSCELOFLANPVTHNHQSAEQHNQDVNSLGLVSAKRTAAVAILLMGSTIYA 473
 QY 538 VLDVNDIRAMEPEFKQFPKXXXLLXGHFGKXXTXXXXELKXKVKKLLKRLQOTNS 597
 DB 474 LCAVDLRLEENLKAQVNCVTVAKVLTTPAGGLHSARFSEKALLTAIDR-EAVYS 532
 QY 598 YDLEP--RWHDAFSXATGTVEXLSXXXXXVSLAANAMKVAAXEKA1--SLTRVYRXX 653
 DB 533 YADDPGSAVYPLMTKIRALVVEHALANGAEKDDSSVFSKITAFEEELREALFREVEAA 592
 QY 654 FKXAPSSSDPALXYL-SPTTRVLYSFVREELGVKARGDVPL-GKQEVTTIGSNVSR1YE 711
 DB 593 RVAFFETAPITRIKESRSFPLVRFVVEELGQ-----VYLGEKLSFGEECNKVFLL 646
 QY 712 IKAGRINXVLVKKL 725
 DB 647 ISERKIDIMLECL 660

RESULT 5

Q8VYG7 PRELIMINARY: PRT; 703 AA.
 AC Q8VYG7:
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCD clade; Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxId=4577;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=cv, Corso;
 RX MEDLINE=9716112;
 RA Rosler J., Krekel F., Amrhein N., Schmid J.;
 RT "Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase
 RT activity";
 RL Plant Physiol. 113:175-179(1997).
 DR EMBL: L77912; AAL40137.1;
 DR GO: GO:0005737; Cytoplasm; IEA.
 DR GO: GO:0016211; F:ammonia ligase activity; IEA.
 DR GO: GO:0016829; F:lyase activity; IEA.
 DR GO: GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
 DR GO: GO:0009058; P:biosynthesis; IEA.
 DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro: IPR008948; L-Asparatase-like.
 DR InterPro: IPR001106; Phe/His NH3lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 703 AA; 74926 MW; 11374FD68516971E CRC64;

Query Match 28.5%; Score 958.5; DB 10; Length 703;
 Best Local Similarity 34.6%; Pred. No. 5.7e-67;
 Matches 255; Conservative 127; Mismatches 242; Indels 113; Gaps 21;

QY 40 GSXLFTYXQTQIDYEXXLDPTDDXKELDGYSLTLGDV--VGAARKGRVYVXSDEL 97
 DB 16 GAAMAEAGSHLDEVKRWVQAQ-ARQPVYKIEGSTRGVAAVSAAXDASGVVLEDEEA 74
 QY 98 RXKIDSVPEPLRXOLXN--SVYGVTTGGSGADRTEDALISQKALLHQLCVLPFSXD 155
 DB 75 RPRVAKSEWILDCIAGGDIYGVTTGPGGSHRRYKQDGAQVLELRHLAGIFGTGD 134
 QY 156 SFYLGRLNSLPLEVVRGAMTRIVNSLTRSHAVRLVLEALTNFLNGITPIVPLRG 215
 DB 135 G-----HTLSEVTRRAAMLVRIINTLQGSQIRFEILATKILNTGVSPLRG 186
 QY 216 ISAGDLSPLVIAAATGHPDSKVHYHHEKEXKIMKREPAIALFGLEP--VLGPYREG 273
 DB 187 ITASGDVPLSYLGLITGRPNAQ-VTVGRK--VDAAEAFKTAGIEGFFFLNPEGL 243
 QY 274 GLVNSTAVSASAKMTALIDHAMLSLSQALITALVEMVGAAGSFHPLHDVTRPHPTQ 333
 DB 244 AIVNGTSVGSALAAVVDANVLAIVLEVSAPCEVWNGKREYTDHLLTKL-KHFGS 302
 QY 334 EVARNIRTLLEGSXFAVHHEEVKDDDEGL--RQDEYPLRTSPQWLGPLVSDIMAHAA 391
 DB 303 EAAALMEHILDGSEFM--KQAKVNELDPLKREKQRYALRTSPQWLGPGI-EVIRA-A 357
 QY 392 VLSLEAG-QSTTDNPLIDVENKXTHHGNGFQASAVXNMTKMTLALIGLNTQUTLTM 450
 DB 358 TKSIEREVSVDNPFVIDVHRGKALHGNQGPFGVSMONALALANIGLMAFQSE 417
 QY 451 LNAQMRLPSCLA-AEDPSLSYHCKGLDPAAAVTSLEGLANPVTHVQPAEMGNQAV 509
 DB 418 VNEFYNGILTNLNGSRNPSLDYGFKGTETAMASYCELDYLNPIPTNHQSADEHNQDV 477
 QY 510 NSLALISARRTAEANVLSLLATHLYCVQAVDLRAMEPEFKQFPKXXXLLXGHFGX 569
 DB 478 NSLGLVSAKRTAEALIDILKMSSTIYALQAVDLRLEENIK----- 520
 QY 570 XXTYXXXXXXELXXVXKXLLKRLQOTNSYDLEPRMDAFSXTATGVEXLSXXXXVAGL 629
 DB 521 -----ASYKNTVTOQAKKVLTMNPSGEL-----SSARFSEKELISA-----IDR 559
 QY 630 AAVNAMKVAAXEKA1SLTRVYRXXF-----FKXAPSSSDP----- 663
 DB 560 EAVFTYEDASASLPLMQYLRALVDHALISSEBERGALRVLDHQVRGAGARGAABG 619
 QY 664 -----ALXYLSPTTRVLYSFVREELGVKARGDVPL-GKQEVTTIGSNVSR1 708

```

Db      620 GGRPRGVAAGTAPFANRIADSGSPFLYFVRBELG-----VLTGEBRLKSPBECNKV 673
QY      709 YEAIKSGRIINXVLYKML 725
      : : : : :
Db      674 FVGISGKLVDEMECL 690

RESULT 6
09AXIS
ID 09AXIS PRELIMINARY: PRT: 711 AA.
AC 09AXIS;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
OS Phorbactis ntl (Violet) (Japanese morning glory).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=35883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Violet.
RA Nakazawa A., Nozue M., Yasuda H., Takeba G., Kubo H.;
RT "Expression of phenylalanine ammonia-lyase in Phorbactis ntl.";
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
EMBL; AF325496; AAG49585.1; -.
DR HSSP; P21310; 1B8F.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia-lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SQ SEQUENCE 711 AA; 77085 MW; CSEGE3A0C0969E90 CRC64;

Query Match 28.4%; Score 955; DB 10; Length 711;
Best Local Similarity 36.3%; Pred. No. 1.1e-66;
Matches 259; Conservative 103; Mismatches 268; Indels 84; Gaps 18;

QY 49 TOLDIYEXXADPXTDDXXELDGYSLTLDVVGAAKRGKXRVVXDSDEIRKXIDKSYEFL 108
      : : : : :
Db 34 SHUDEKXWVAE-FRKPAYKLGEGITLVAQVAIAISRDNATVVELSESRAGVASSDWM 92
      : : : : :
QY 109 RQOLXNSV--YGVTTGFGSGADTRTEDAISLQKALLEHQLCGVLEFTSXDSFXLGRGLE-- 164
      : : : : :
Db 93 MDSMNKGTDSYGVTTGFGATSHRRTKOGALQKELIRFLNAGI-----FANGTESC 143
      : : : : :
QY 165 NSLPLEVVGAMTIRNSLIRGHSAVELVLEALNPNHGITPIVPLRGITASGDISP 224
      : : : : :
Db 144 HTTPHSATPAAMIVRNITLLQSGISGIRFELPAITKLNNHNTIPCLPRGHTTSGDILVP 203
      : : : : :
QY 225 LSYIAAATGHPDSKXVHXHEGEXKIMAREALFGLER--VVLGPRGGLVNGTAVS 282
      : : : : :
Db 204 LSTIAGLIGRPNKX-VGPNG--ETLNAEALRLAGVVGFFELQPRGGLAVNGTAVG 260
      : : : : :
QY 283 ASNATLALHDAMLSLISQALTLATVEAMVGAAGSFHPLDVTREPHTQIEVANIRTL 342
      : : : : :
Db 261 SGASAVLFEANLAVLSEVLSAIFAEVNNGRP-FETDHLTHKLGHGQIEAALMEHI 319
      : : : : :
QY 343 LSGSFAVHEHEEVKCKDEGLIADODRYPLRTSPOMLGLIVDSIHAAVAVLSLAGOSTT 402
      : : : : :
Db 320 LDSSSY-VRAAQGCHMDPLQKPKORVALKRSPPWLGQI-EVIRATKXIHREINSVN 377
      : : : : :
QY 403 DNPDLIVENKXTHGNGFQASAVXNTMEKTRIALALIGKLNFTQITLMLNAGNGLPSC 462
      : : : : :

```

```

Db      378 DNPDLIDVARSKALHGNGFQGTPIGVSMDSNRLALASIGKLLFAQFSELVANDYNNGLPSN 437
QY      463 L-AAEDBLSLHCGGLDIAAAYTSELGHLANPYTHVQAEKNGNQAVNSLALIASARTA 521
      : : : : :
Db      438 LTAGRNPBLDYGFGAELMASYCSLEQLANPYTNVQSBCHQNDVNSLGLISARKTA 497
      : : : : :
QY      522 EANDVLSLLATHYCVLQAVDLRAMEFEFKQCFXPXXXXLXQ-----H----- 566
      : : : : :
Db      498 EAVDYLKIMSSTYVALCQALDLRFLLENNLRNAVNAVTOVAKRTLTMGANGELHPARFC 557
      : : : : :
QY      567 -----FGXXTXXXXXXELKXKXKXKXRLQRTNSYDLPEPMHDAFSATG 613
      : : : : :
Db      558 EKDLRLVVDREVFAVADDPGSANYPLMQKLRQALVDHALQNGSESKN-----TG 607
      : : : : :
QY      614 TVEKLSXXXXXXVSLAAVNAWKAAEKAISLTRXYRXXFWXAPSSSPAL--XYLSPR 671
      : : : : :
Db      608 TSI-----FLKVAFADELKAVLPEVEAR-----IAGESGNALIPRIEGR 651
      : : : : :
QY      672 TRVLYSFYREELGVKARGDYFLQKQEVTTGSNVSRITYEAIKSGRIINXVLYKML 725
      : : : : :
Db      652 SYPLKYFVREGIGT-----ELTGEKVRSPBECDKVFTAMCEGSIIDPLECL 700

RESULT 7
09M567
ID 09M567 PRELIMINARY: PRT: 730 AA.
AC 09M567;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase 2 (EC 4.3.1.5).
GN PAL2.
OS Rubus idaeus (Raspberry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucotyledons; Rosales; Rosaceae; Rosoideae; Rubus.
OX NCBI_TaxID=32247;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21437948; PubMed=11553751;
RA Kumar A., Ellis B.E.;
RT "The phenylalanine ammonia-lyase gene family in raspberry: structure,
RT expression, and evolution.";
RL Plant Physiol. 127:230-239(2001).
EMBL; AF237955; AAF40224.1; -.
DR HSSP; P21310; 1B8F.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia-lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro; IPR008948; L-Aspartase-like.
DR InterPro; IPR001106; Phe/His NH3lyase.
DR InterPro; IPR005922; Phe_am_lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01226; phe_am_lyase; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase.
SQ SEQUENCE 730 AA; 79422 MW; 0B542417B50CE113 CRC64;

Query Match 28.4%; Score 955; DB 10; Length 730;
Best Local Similarity 36.8%; Pred. No. 1.1e-66;
Matches 257; Conservative 107; Mismatches 269; Indels 52; Gaps 19;

QY 40 GSXLPXTXQTOLDIYEXXADPXTDDXXELDGYSLTLDVVGAAKRGKXRVVXDSDEIRX 99
      : : : : :
Db 44 GAABSMGTSIHDEVRVVAE-YRKPVYKLGEGITLTSQVAIAINHDSGVVLEASARA 102
      : : : : :
QY 100 KIDKSYEFLRQOLXNSV--YGVTTGFGSGADTRTEDAISLQKALLEHQLCGVLEFTSXDSF 157
      : : : : :
Db 103 GYKASSDWMDSMNKGTDSYGVTTGFGATSHRRTKOGALQKELIRFLNAGVLANGTES- 161
      : : : : :

```

```

QY 158 KLGRGLNSLPLEVRGAMTIRVNSLTRGSAVRLVVALTNPLNHTPTPLVPLRGTTIS 217
D 162 -----AHLPHSATRAAMLVRLNTLQSGYSGRFLBIAIKFNHNHTPCLPLRGTTIT 215
QY 218 ASGDLSPSYIAAATGHPDSKVHHEGKEMKAREIAIFGLEP--VVGPRGGL 275
D 216 ASGDLVPSYIAGILTRPNSKA-VGPKG--ETLNAAEFAQVIGSSGFELQPRGGL 272
QY 276 VNGTAVASAMATLALDHAKLSLSQALTALTYEANVAGSGFPHLDVTRPHPTQIEV 335
D 273 VNGTAVAGSGLASTLPETNLTALITSLTSAIFAEVVGQKP-EFTDHLTKLGHHPQIEA 331
QY 336 ARNRTLLGSSXFVHHEEVKVKDEGLRQDRYPLTSPQMLGPIVSDMTHAVALSL 395
D 332 AATMEHLIDGSSY-VKAEKTLHQDPLQKPKDRYALRTSPQLGPOIE--VIRSTSTSI 388
QY 396 EAG-QSTTNDPLIDVENKXTHGNGFQASAVXNTVEKTRIALALIGKNTQTLEMLNAG 454
D 389 EREINSVNDPLIDVSHNKALHGGNFQGTPIGVSMNTRLAISIGKLMFAQFSELVNDP 448
QY 455 MNRLSPSCA-AEDPSLSTCKGLDRAAAYTSELGHLANPTTVVQPRMGNQAVNSLA 513
D 449 YNNGLPSNLGSGRPSLDYGFKAELIAMSAYCELOFLANPVTNVQSAEQHNDVNSLG 508
QY 514 LISARTRAEANDVLSLLATLHYCVLAVDLRAMEFEFKQFPXKXLLXQHPGXKXTX 573
D 509 LISRKTAEAVDILKLMSSFLVALQALDLHLEENLKSJTKNTVSQLAKVLTITVNG 568
QY 574 XXXXXLXKXKXKXKLEQINSYDLPRMHDAPSXATGVVEKLSXXXXXVSLAAN 633
D 569 ELHSPRCEDLMLVVER-EYLPAYIDP-----CSATYPLMQGLRQ--VLVEHALTN 618
QY 634 A-----MKVAAEKAS--LTRVYRXFXWAPSSSPAL--XLSPTRVLYSF 678
D 619 GENEKASTIFOKITAFEBELKTLIPKVEESA-RAIYSGNAALPNRIVECRSPLYKF 677
QY 679 VREELGVKARGDVFQKQEVITGSNVSRITYAISKGR 717
D 678 VREELG-----GEPLTGEKVSFGEECDKVFAMCQGN 711

RESULT 8
Q94ENO PRELIMINARY; PRT; 708 AA.
AC Q94ENO;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN PALI.
OS Rhamania glutinosa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Lamiales; Lamiales; incertae sedis; Rhamania.
OX NCBI_TaxID=99300;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Keunshan;
RA Yun S.J., Lee B.K., Park M.R., Srinivas B., Chun J.-C.;
RT "Molecular characterization of phenylalanine ammonia-lyase gene in
RT Rhamania glutinosa.";
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR GO:0005737; C:cytoplasm; IEA.
DR GO:0016211; F:ammonia lyase activity; IEA.
DR GO:0016229; F:lyase activity; IEA.
DR GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
DR GO:0009058; P:biosynthesis; IEA.
DR GO:0006559; P:phenylalanine catabolism; IEA.
DR InterPro: IPR008948; L:aspartase-like.
DR InterPro: IPR001106; L:aspartase-like.
DR InterPro: IPR005922; Phe.am_lyase.
DR InterPro: IPR006162; Pplantn_S.

```

```

DR Pfam; PF00221; PALI.
DR TIGRfams; TIGR01226; phe.am_lyase.1.
DR PROSITE; PS00488; PALI_HISTIDASE.1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE.1.
KM Lyase.
SQ SEQUENCE 708 AA; 76835 MW; 1B6AF08816F1B8FE CRC64;

Query Match 28.4%; Score 954; DB 10; Length 708;
Best Local Similarity 36.7%; Pred. No.1.3e-66;
Matches 253; Conservative 103; Mismatches 260; Indels 74; Gaps 18;

QY 68 ELDGYSLTGVDVGAARGRVYRVDSPEIKTKDKSVEFLRXQJXNSV--YGTPTGFG 125
D 50 KIGGESLTIAQVAAIAAADNVAVALATAAGVKAASDWMEMKGTSTYGTTFGGA 109
QY 126 SADRTEDAIISLQKALHQLQGLVFTSXDSFXLGRGLNSLPLEVRGAMTIRVNSLTR 185
D 110 TSHRRTQGGALQKELIFFLNAGIFNGTSS-----NHALPHSATRAAMLVRLNTLQ 162
QY 186 GHSAYLVVLEALTMFLNHGTPVPLRGTSASGDLSPSYIAAATGHPDSKVHXYHE 245
D 163 GTSGLRFTLELTLFPLNHNITPCLPLRGTTASGDLVPSYIAGILTRPNSKA-VGPN 221
QY 246 GXEKIMXAREALTLGLEPVP-LGPKESGLVNGTAVASAMATLALDHAKLSLSQALT 304
D 222 G--EALNGEARSLAGVGGFELQKESGLAVNGTAVGSGLASIALVDANILAVSEVTS 279
QY 305 ALTVEAMGHASGFPPLHDVTRPHPTQIEVANIRTLISGSXFAVHHEEVKVKDEGI 364
D 280 VIFAEVMNGKP-EFTDHLTKLKHHPQIEAAMEHLIDGSAV-VKAEKTLHETDPLQK 337
QY 365 LRQDRYPLTSPQMLGPIVSDMTHAVALSLAEGQSTNDPLIDVENKXTHGNGFQASA 424
D 338 PHQDRYALRTSPQMLGPIEIVIRATKIEBEI-NSVNDPLIDVSNKALHGGNFQGT 396
QY 425 VANTVEKTRIALALIGKNTQTLEMLNAGNRLPSCLA-AEDPSLSTCKGLDIAAAA 483
D 397 IGVSMDNTRLIAALIGKLMFAQFSELVNDPENNGLPSNLGGRPSLDYFKSEIAMS 456
QY 484 YTSSELGHLANPTTVVQPRMGNQAVNSLISARTRAEANDVLSLLATLHYLCVLAQVD 543
D 457 YCELOFLANPVTNVQSAEQHNDVNSLGILISRKTVEALDILKLMSSFTYLALCOAVD 516
QY 544 LRAME-----FEFKQFPXKXLLXQHF-GXXXTXKXKXELKXKXKX----- 587
D 517 LHLLENRLSVKNTVSOVAKRTLTMWINGELHPSRCEKDLRVDRVYFAYIDPCS 576
QY 588 -----LXKRLQCT-----NSVDLEPRMHDAPSXATGVVEKLSXXXXXVSLAANAKRY 637
D 577 GTFPLMQKLRQVLDVHDALNNGSEKN---VSTISFQKIAFE-----VELXAILPKVEY 626
QY 638 AAXAEKALSLTRVXKXFXWAPSSSPAL--XLSPTRVLYSFVREELGVKARGDVFQ 695
D 627 ESARIAL-----ESGNPAIGRNITECRSPYLYKFIRELGT-----NYLTG 667
QY 696 KOEVITGSNVSRITYAISKGRINXVLYKML 725
D 668 EKVSFGEECDKVFALSKGLIVDPLKCL 697

RESULT 9
Q8RUZ3 PRELIMINARY; PRT; 661 AA.
AC Q8RUZ3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Phenylalanine ammonia-lyase (EC 4.3.1.5) (fragment).
GN PALI.
OS Pinus sylvestris (Scots pine).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=33349;

```

QY		481	AAAYSEIGHLNPPTTHVOPKPMGNQAVNSALISARTAEADVLSILATHTLVQLQ	516
Dd		457	NMSITSELINLPNVPISHVQSABQHNVDSUGLVARSKAFAIDIKMLDSTLYTLRCLQ	516
QY		541	AVDLRAWE 548	
Dd		517	AVDLRLHE 524	
 RESULT 10				
ID	Q852S1	PRELIMINARY,	PRT,	715 AA.
AC	Q852S1;			
DT	01-JUN-2003 (Tremblrel, 24, Created)			
DT	01-JUN-2003 (Tremblrel, 24, last sequence update)			
DT	01-OCT-2003 (Tremblrel, 25, last annotation update)			
DE	Phenylalanine ammonia-lyase (EC 4.3.1.5).			
CN	GPCPAL3.			
OS	Daucus carota (Carrot).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;			
OC	Campanulales; Apiales; Apiaceae; Apioidae; Scandiceae; Daucinae;			
OC	Daucus.			
OX	NCBI_TaxID=4039;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv_Kurodagosun;			
RA	Ozeki Y., Chikagawa Y., Kimura S., Sch H., Maeda K., Pormsirirong W.,			
RA	Kato M., Akimochi H., Oyagai M., Fukuda T., Koda T., Itoh Y.,			
RA	Yamada A., Ueno H., Takeda J.,			
RT	"Putative cis-elements in the promoter region of phenylalanine			
RT	ammonia-lyase gene of carrot involved in anthocyanin biosynthetic			
RT	pathway."			
RU	Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AB089813; BAC56977.1; -.			
DR	GO; GO:0005737; Cytoplasm; IEA.			
DR	GO; GO:0016211; Gamma-ligase activity; IEA.			
DR	GO; GO:0045448; F-phenylalanine ammonia-lyase activity; IEA.			
DR	GO; GO:0009058; P-biosynthesis; IEA.			
DR	GO; GO:0006559; Phenylalanine catabolism; IEA.			
DR	Interpro; IPR008948; L-Asparaginase-like.			
DR	Interpro; IPR001106; Phe/His NM3-Lyase.			
DR	Refam; PF00221; PAL; 1.			
DR	TIGRFAMs; TIGR01226; phe_am_lyase; 1.			
DR	PROSITE; PS00488; PAL_HISTIDASE; 1.			
KM	LYase.			
SC	SEQUENCE 715 AA; 77802 MW; 8006779SD5756584 CRC64;			
 Query Match 28.1%; Score 944.5; DB 10; Length 715; Best Local Similarity 36.0%; Pred.No. 7.5e+66; Matches 251; Conservative 111; Mismatches 250; Indels 85; Caps 194				
QY		68	ELDGYSLTIGDVVS-AARKGRVRYXDSDELRIKXIDSVEFLRKQLNSV--VGVTGPG	1244
Dd		54	KUGGEFLTISYAALSAISARDSDGVKEVLEEARAGKASSDWMSKNKGTDSTGGVTGFG	1133
QY		125	GSADTTEPAISLIQTALLLEHOLCGVLPTPSXDSPFXLGGLG--NSLPLEVRGMATIRVN	1811
Dd		114	ATSHRRTRKGGALQRELRFRFNAGI-----FGSGAEAKGNNTLPHSATRAAMCVRIN	1644
QY		182	SLTRGHSARLVVLEALNFNLNHGTTPVPLRGTTASGDLSPLSYIAAIIATHPDSKYH	2411
Dd		165	TLTGQVSGIRFEILAIKFKPLNNHTTPELPRGTTTASGDIVPLSYIAGLLTGRPNKA-	2233
QY		242	VHHEGXEKIMKAREALFGLER--VVVGPRGGGLVNGTAVASMAVTALAHDAHMLSL	2299
Dd		224	VGPTGV--TLSPBEAFKLAVGEGGFELLQPKRGIALVNGTAVGSMASVLFENATLAVL	2811
QY		300	SOALTPLYEAMVWGAGSFHPLEDVTRPHPTQLEVARNIRTLLEGXFPVHNHEEVKKV	3599

QY 360 DDEGILRQDRYPLRTSPQWLGPIVSDMIAHAVALSLIAGOSTTNDPLIDVENKXTHGNG 419
 Db 340 DPLQKPKQDRYALRTSPQWLGPIVSDMIAHAVALSLIAGOSTTNDPLIDVENKXTHGNG 398
 QY 420 FQASAVNTEKTRTALALIGKLNFTQLEMLNAGNNRGLPSCLA-AEDPSLSYCKGLD 478
 Db 399 FQGPFGVSMNDRTRTALALIGKLNFTQLEMLNAGNNRGLPSCLA-AEDPSLSYCKGLD 458
 QY 479 TAAAYTSELGHLANPTTHVQPAEMNGQAVNSALISARTRAEANDVLSLLATHLVCV 538
 Db 459 TAAAYTSELGHLANPTTHVQPAEMNGQAVNSALISARTRAEANDVLSLLATHLVCV 518
 QY 539 LQAVDLRAMEPEFKQCPKXXXLLKQHFQXXXLXXXXXXELXXKX----- 584
 Db 519 CQALIDLRHLEENK-----STVNTVSVQAKRVLTVNGVNGELHPSRF 560
 QY 585 -XXIXKRLBQ-----TNSYDLERPMHDAFSXATGVVEXLSXXXXXVSLAAV 632
 Db 561 CENKLLRVDBREYFAIYDPCGATYPLMCKLE-----TLVEHALNNGDKERNLST 613
 QY 633 NAKVVAAYAEKAIIS--LTRXVXXFWXAPSSSSPAL--XVLSPTRVLSFVFEELGVGAR 688
 Db 614 IFQKIAAFEBDELKALLPEKEVESA-RAVESGNPAIPNRIRKCRSYPLVYFVEBELGT--- 669
 QY 689 RGVDFLGKQEVITGSNVSRITYEAIKSGRINXVYVRL 725
 Db 670 --EYLTEBKVTSPEEFEDKVFYTMTKGEIIDLPLECL 704

RESULT 11

Q8AS6 PRELIMINARY; PRT; 661 AA.

AC Q8AS6: 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
 GN PAL.
 OS Pinus sylvestris (score pine).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophytes; Coniferopsida; Coniferales; Pinaceae; Pinus.
 OC NCBI_TaxID=3349;
 RN NCBI_TaxID=3349;
 RP SEQUENCE FROM N.A.
 RC STRAIN=K130;
 RX MEDLINE=21660210; PubMed=11801746;
 RA Dvornyk V., Sirovko A., Mikkonen M., Savolainen O.;
 RT "Low Nucleotide Diversity at the pal1 Locus in the Widely Distributed
 RT Pinus sylvestris.";
 RL Mol. Biol. Evol. 19:179-188 (2002).
 DR EMBL: AF353981; AAL74311.1; -.
 DR GO: GO:0005737; C:cytoplasm; IEA.
 DR GO: GO:0016211; F:ammonia-lyase activity; IEA.
 DR GO: GO:0016829; F:lyase activity; IEA.
 DR GO: GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
 DR GO: GO:0009058; P:biosynthesis; IEA.
 DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro: IPR008948; L-Asparagine-lyase.
 DR InterPro: IPR001106; Phe/His NH3-lyase.
 DR InterPro: IPR005922; Phe/His NH3-lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 FT NON TER 1 1
 FT NON TER 681 681
 SQ SEQUENCE 681 AA; 74341 MW; 5BFCC373969FFBD0 CRC64;

Query Match 28.0%; Score 943.5; DB 10; Length 681;
 Best Local Similarity 43.9%; Pred. No. 8,4e-66;
 Matches 214; Conservative 85; Mismatches 162; Indels 27; Gaps 10;

QY 69 LDGYSITLDVYGAAKRGKXVAVXSDIEIRKKIDKXVEFLKQX--NSVYGYTTFGGGS 126
 Db 56 IEKSLTIDVAVVARRGQVKYKLDIAAKRREVSNNWVLTQMTKTDTTGVTTGGAT 115
 QY 127 ADTRTEDALSLQKALLEHQLCGVLPYTSXDSFYLGRGLENLPLEYVNGMTIRVNSLIRG 186
 Db 116 SHRTNQGAELOKELLIRLNGV-----LGKCEPNVASEDTTAAALVTRNTLLQG 166
 QY 187 HSAVRLVIEBALTNPLNHSITPIVLRGTISASGDSLSTIAALIGHPKSKHYVHEG 246
 Db 167 YSGIEMWDLIEVEKELNGLTPKPLRGITIASGDLVLSYAGILTGRPSRVV-SRDG 225
 QY 247 XEKIKKARBAIALFGLF-PVVLGPKKGGGLVNGTVASAMATLALHPAHLISLSQALTA 305
 Db 226 IF--NSGAKALKVGLERKLPFLOPEGLATNGNSVGAALSLTICPDANVALLSEVISA 283
 QY 306 LTVANVGHAGSFHPLHDVTRPHPTQIEVARNITLIEGSXFVH---HEEVKXKXD 361
 Db 284 MFCENVNGKPEFTDPLTKL-KLHPGQMEAAIMBYVDGSSYMKHAKLHNNPLQKP- 341
 QY 362 EGLTRQDRYPLRTSPQWLGPIVSDMIAHAVALSLIAGOSTTNDPLIDVENKXTHGNG 421
 Db 342 ---KQDRYALRTSPQWLGPIVSDMIAHAVALSLIAGOSTTNDPLIDVENKXTHGNG 396
 QY 422 ASAVNTEKTRTALALIGKLNFTQLEMLNAGNNRGLPSCLA-AEDPSLSYCKGLDIA 480
 Db 397 GTPFGVSMNDRTRTALALIGKLNFTQLEMLNAGNNRGLPSCLA-AEDPSLSYCKGLDIA 456
 QY 481 AAAYTSELGHLANPTTHVQPAEMNGQAVNSALISARTRAEANDVLSLLATHLVCV 540
 Db 457 MASYTELLYLANPTTHVQPAEMNGQAVNSALISARTRAEANDVLSLLATHLVCV 516
 QY 541 AVDLRAME 548
 Db 517 AVDLRAME 524

RESULT 12

Q9MAX1 PRELIMINARY; PRT; 716 AA.

AC Q9MAX1: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase.
 OS Catharanthus roseus (Rosa periwinkle) (Madagascar periwinkle).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Gentianales; Apocynaceae; Rauvolfioideae; Vincaceae;
 OC Catharanthus.
 ON NCBI_TaxID=4058;
 RN NCBI_TaxID=4058;
 RP SEQUENCE FROM N.A.
 RA Kiyoca S., Sakano K., Kim U.;
 RT "cDNA of phenylalanine ammonia-lyase from Catharanthus roseus.";
 RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB042520; BA95629.1; -.
 DR HSP: P21317; 188F.
 DR GO: GO:0005737; C:cytoplasm; IEA.
 DR GO: GO:0016211; F:ammonia-lyase activity; IEA.
 DR GO: GO:0016841; F:ammonia-lyase activity; IEA.
 DR GO: GO:0016829; F:lyase activity; IEA.
 DR GO: GO:0009058; P:biosynthesis; IEA.
 DR GO: GO:0006559; P:phenylalanine catabolism; IEA.
 DR InterPro: IPR008948; L-Asparagine-lyase.
 DR InterPro: IPR001106; Phe/His NH3-lyase.
 DR InterPro: IPR005922; Phe/His NH3-lyase.
 DR Pfam: PF00221; PAL; 1.
 DR TIGRFAMs: TIGR01226; phe_am_lyase; 1.
 DR PROSITE: PS00488; PAL_HISTIDASE; 1.
 KW Lyase.
 SQ SEQUENCE 716 AA; 78241 MW; 1F6D4DB8683B0B CRC64;

Query Match 28.0%; Score 942.5; DB 10; Length 716;
 Best Local Similarity 35.7%; Pred. No. 1.1e-65;
 Matches 256; Conservative 120; Mismatches 270; Indels 71; Gaps 22;

QY 40 GSKLPTTXXTOADIYEXXADXTDDXXELDGYSTLTGLDVGW-AARKGRYRVXDSEIR 98
 Db 29 GMAADSLKSHLDEKRYVAE-FRPRVYKLGSETITISQVAAIAARDNAVKVLSSEDA 87
 QY 99 XKIDKVEFLRXQLNSV--YGVTTGFGSADRTEDAISLOKALLEHQLGVLPTSXDS 156
 Db 88 AGVKASDVMWESMNKGTDSYGVTTGFGATSHRRTQCGALQKELIRFLNAGI----- 140
 QY 157 FXKGRLEMS--LPLEVVGAMTIRVNSITRSHSVRLVLEALTNPLNPTITVPLRG 214
 Db 141 --FNGTSSHTLPHSATPAALVRLNITLQSGIRSEIRLEALTKFLNHTPLCLPBG 198
 QY 215 TISASGDLPSLYAAITGHPDCKVYHHEGKIMKAREALFGLPEVY--LGPREG 272
 Db 199 TITASGDLVPLSYIAGLTGPRNSKA-VGPNG--EIVNPEQAFKXAGVNDGLFELQREG 255
 QY 273 LGIVNTAVSAGMATLADHMLLSLQALTALVEMVNGHAGSFHFDVTRPHPTQ 332
 Db 256 LALVNTAVGSGASVWLEFANILVLSSEVLSAIEAVNNGRP-EFTDHLTKLKHHPGQ 314
 QY 333 IEVARNIRTLGSGXF--AVH--HEEVVKYKDEGILRODRYPLRTSPQMLGPLVSMIH 388
 Db 315 IEAAATHEHILDSGVYKAHKLHENDPQRP-----KQDRYALRTSPQMLGPLVIRS 369
 QY 389 AAHVISLEAGQSTTDNPLIDVENKXTHHGNFQASAVXNTWEKTRLALIGKINTQLT 448
 Db 370 ATQCIEREI--NSVNDPLIDVSRNKALHGNFQGPFGVSMNARLALASIGKMFQIS 428
 QY 449 EMNNAEMNGLSCLA-AEDPSLSYHCKGLDIAAAYTSEIGHANPTTHVQPAEENGQ 507
 Db 429 ELVNDYRNGLSNLSGKNPISLVYGFKAETIAYASTCELOYLANPTTHVQSAEQHNQ 488
 QY 508 AVNSIALISARTAEANDVLSLLATHLVYCLQAVDLAMEEFPKQFPXXXXLXQHF 567
 Db 489 DVNSIGLISRTKAEAVETILKMSYVALCOALDLRHEENLKNVAKNTVQVAKRTL 548
 QY 568 -----GXKTXKXXXXXELKXKXKXKLEGTNSYDLEPRKHAFSAKGT----- 614
 Db 549 TVGVNDELHPSRCEKDLIRVDR-----EYVFAVVDP-----CSGYPLMEKLR 594
 QY 615 --VVEKLSSXXXXXVSLAAVNANKVAXAEKAS--LTRXVXXFXWAPSSSPAL--XYL 668
 Db 595 QYIVDAHALONGSEKKNVNTSIFOKIAAFEDELKTYLPRKVESR-RTALENGNPALPRIK 653
 QY 669 SPTTRVLYFVREBELGVKARGDVLGQCVTTIGSNVSRITFAIKSGRIKXVLYMVL 725
 Db 654 ECRSYPLVYFVEDVG-----AEFLGEBKDRSPGEEFDPKVFAMCNKRTIDPLECL 705

RESULT 13
 ID QBRV49 PRELIMINARY; PRT; 681 AA.
 AC QBRV49;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase (EC 4.3.1.5) (Fragment).
 GN PAL1.
 OS Pinus sylvestris (Scots pine).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferophyta; Coniferales; Pinaceae; Pinus.
 CX NCBI_TaxID=3349;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K013, and K014;
 RA MEDLINE=21660210; PubMed=11801746;
 RA Dvoynik V., Sirovko A., Mikkonen M., Savolainen O.;
 RT "Low Nucleotide Diversity at the pal1 Locus in the Widely Distributed
 Pinus sylvestris.";

RL Mol. Biol. Evol. 19:179-188 (2002).
 DR EMBL; AF353974; AAL74324.1; -
 DR EMBL; AF353975; AAL74325.1; -
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0016211; F:ammonia ligase activity; IEA.
 DR GO; GO:0016829; F:lyase activity; IEA.
 DR GO; GO:0045548; F:phenylalanine ammonia-lyase activity; IEA.
 DR GO; GO:0003058; F:biosynthesis; IEA.
 DR GO; GO:006559; F:phenylalanine catabolism; IEA.
 DR InterPro; IPR008948; L-Asparagine-like.
 DR InterPro; IPR001106; Phe/His NH3lyase.
 DR InterPro; IPR005922; Phe_am_lyase.
 DR Pfam; PF00221; PAL; 1.
 DR TIGRfam; TIGR01226; phe_am_lyase; 1.
 DR PROSITE; PS00488; PAL_HISTIDASE; 1.
 KW lyase.
 FT NON TER 1
 FT NON TER 681
 SQ SEQUENCE 681 AA; 74394 MW; 35BBA50BA8480286 CRC64;

Query Match 28.0%; Score 941.5; DB 10; Length 681;
 Best Local Similarity 43.9%; Pred. No. 1.2e-65;
 Matches 214; Conservative 84; Mismatches 163; Indels 27; Gaps 10;

QY 69 LDGSLTLGDVVGAAKGRXRVXDSDEIRKXIDKSYVEFLRXQLX--NSYGVTTGFGGS 126
 Db 56 IEGKSLTISDVAAVARRQGVKVLDEAKSRVESSKWTLTQMTKGTDTGVTGFGAT 115
 QY 127 ADRTEDALSLOKALLEHQLGVLPTSXDSFXLGRGLNLSPLEVVGAMTIRVNSITRG 186
 Db 116 SHRRNOCALQKELIRFLNAGV-----LKCPENVLSBPTTAAMVRLNITLLOG 166
 QY 187 HSAVPLVLEALTNPLNNGITPIVPLRGTISAGPLSPLSYIAAITGHPDCKVYHHEG 246
 Db 167 YGIRMDILEYVEKLLNGLTPKLPRLRTITASGDLVPLSYIAGLTGPRNSRR-GRDG 225
 QY 247 XEKIKXARBAALFGLF-PVVLGPREGGLVNGTAVASAVATLADHMLLSLQALT 305
 Db 226 IE--MSGAEALKQVLEKRFELQPKEGIAIVNGTSVGAALASIVCPANVALISEVISA 283
 QY 306 LTVEMVNGHAGSFHFDVTRPHPTQIEVARNIRTLGSGXFVH--HEEVVKYKD 361
 Db 284 MFCVWVNGKPEPTDPLTKL--RHPQMEALAIWEYLDGSSYKKAALKHNNPQRP- 341
 QY 362 EGLRODRYPLRTSPQMLGPLVSMIHAAVLSLQASTDNPIDVENKXTHHGNFQ 421
 Db 342 ----KQDRYALRTSPQMLGPLVSMIHAAVLSLQASTDNPIDVENKXTHHGNFQ 396
 QY 422 ASAAXNTWEKTRLALIGKINTQLTMLNAGNRGLPSCLA-AEDPSLSYHCKGLDIA 480
 Db 397 GPPIGSMNRLSLISAIGKMFQAFSELVNDYNGGLPSNLGGPPPSLDYGKGAIEA 456
 QY 481 AAAYTSEIGHANPTTHVQPAEENGQAVNSIALISARTAEANDVLSLLATHLVYCLQ 540
 Db 457 MASYSSELLYLANPTSHVQSAECHOVDNSLGLVSARKSAEALDILKMLSTYTLALCQ 516
 QY 541 AVDLAME 548
 Db 517 AVDLRLIE 524

RESULT 14
 ID QBRW4 PRELIMINARY; PRT; 717 AA.
 AC QBRW4;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Phenylalanine ammonia-lyase.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;


```

Db      181 NSLTGSHGAVLVLEALTNFLNHTGIVPLRGTSASGDLSPSYIAAATGHPDSKV 240
QY      241 HXHEGKEXIKAKARALIFGLPEVYLGPKKGLGVNQTAAVSAMATLALHDAMLSLS 300
Db      241 HXHEGKEXIKAKARALIFGLPEVYLGPKKGLGVNQTAAVSAMATLALHDAMLSLS 300
QY      301 QALTALTYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
Db      301 QALTALTYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
QY      361 DEGLIRQDRYPLRTSPQWLGPLVSDMTIAHAVALSLEAGOSTTNDPLIDVENKXTHHGNF 420
Db      361 DEGLIRQDRYPLRTSPQWLGPLVSDMTIAHAVALSLEAGOSTTNDPLIDVENKXTHHGNF 420
QY      421 QASAVNTEKTRIALALIGKLNFTQLTETMLNAGNRLPSCLAEDPSLSYHCKGIDIA 480
Db      421 QASAVNTEKTRIALALIGKLNFTQLTETMLNAGNRLPSCLAEDPSLSYHCKGIDIA 480
QY      481 AAAYTSELGHLANPYTTHVQPAEMGNQAVNSLALISARRTEANDVLSLLATHTLYCVLQ 540
Db      481 AAAYTSELGHLANPYTTHVQPAEMGNQAVNSLALISARRTEANDVLSLLATHTLYCVLQ 540
QY      541 AVDLRAMEFEFKQRPXKXXLLXQHPGXXTXKXXKXXKXXKXXKXRLBQTNSTYDL 600
Db      541 AVDLRAMEFEFKQRPXKXXLLXQHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 600
QY      601 EPRMHDAPFSGATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRVXXKFWXAPSS 660
Db      601 EPRMHDAPFSGATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRVXXKFWXAPSS 660
QY      661 SSPALXYLSPTRTVLYSPVREELGVKARGDVFLEKQVTTGNSVSRLEYAKSGRLNXY 720
Db      661 SSPALXYLSPTRTVLYSPVREELGVKARGDVFLEKQVTTGNSVSRLEYAKSGRLNXY 720
QY      721 LVKMLA 726
Db      721 LVKMLA 726

RESULT 3
AAP80513
ID AAP80513 standard; protein; 716 AA.
XX
AC AAP80513;
XX
DT 25-MAR-2003 (revised)
DT 12-NOV-1990 (first entry)
XX
L-phenylalanine ammonialyase.
XX
L-phenylalanine ammonialyase; PAL; foreign gene expression;
XX
culture temperature; expression regulation.
XX
Rhodospiridium toruloides.
XX
EP279665-A.
XX
PD 24-AUG-1988.
XX
PF 18-FEB-1988; 88EP-00301356.
XX
PR 19-FEB-1987; 87JP-00034397.
XX
PR 18-JUN-1987; 87JP-00152359.
XX
PA (MITK ) MITSUI TOATSU CHEM INC.
XX
PA Fukuhara N, Yoshino S, Yamamoto K, Sone S, Suzuki M, Nakajima Y,
XX
DR WPI; 1988-236895/34.
XX
DR N-PDB; AAN81116.
XX
PT Regulation of expression of foreign gene in Escherichia coli - by

```

```

PT      maintaining temp. of culture at 40 deg. C or more to suppress expression.
XX
XX      Claim 3; Page 15-19; 36pp; English.
CC      The PAL gene is expressed in E.coli carrying a recombinant plasmid.
CC      Expression is regulated by maintaining the temperature at at least 40
CC      deg. C to suppress expression. Pal is produced in high concentrations.
CC      See also EP-279664; regulation by sugar component as C source at 0.3% or
CC      more. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-
XX      2003 to correct PI field.)
XX
SQ      Sequence 716 AA;
XX
Query Match      89.8%; Score 3021; DB 1; Length 716;
Best Local Similarity 84.0%; Pred. No. 4.5e-300;
Matches 610; Conserved 25; Mismatches 81; Indels 10; Gaps 3;
QY      1 MAPSDSIATSGXANGXNKGHAAXXASXXXXXXXASXKPTXXTOLDIVEXXLD 60
Db      1 MAPSDSIATSGXANGXNKGHAAXXASXXXXXXXASXKPTXXTOLDIVEXXLD 60
QY      61 PXTDXEELDGYSLTLDGVDVGAARKGXVRYXSDERKXKIDKSEFLRXQLXNSVYGT 120
Db      55 P-TDSTLELDGYSINLGDVVAARKGRPVRYKDSDEIRSKIDKSEFLRQLSWSYGV 113
QY      121 TGFSSADTRFEDALSLQKALLEHQLGCVLPTXSDXFLGRGENSLPLEVVGAMTIRV 180
Db      121 TGFSSADTRFEDALSLQKALLEHQLGCVLPTXSDXFLGRGENSLPLEVVGAMTIRV 180
QY      114 TGFSSADTRFEDALSLQKALLEHQLGCVLPTXSDXFLGRGENSLPLEVVGAMTIRV 173
Db      181 NSLTGSHGAVLVLEALTNFLNHTGIVPLRGTSASGDLSPSYIAAATGHPDSKV 240
QY      174 NSLTGSHGAVLVLEALTNFLNHTGIVPLRGTSASGDLSPSYIAAATGHPDSKV 233
Db      241 HXHEGKEXIKAKARALIFGLPEVYLGPKKGLGVNQTAAVSAMATLALHDAMLSLS 300
QY      234 HYVHEGKEXIKAKARALIFGLPEVYLGPKKGLGVNQTAAVSAMATLALHDAMLSLS 293
Db      301 QALTALTYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 360
QY      294 QALTALTYEAMVGHAGSPHPLHDVTRPHPTQIEVARNIRTLGSGXFAVHHEEVKVD 353
Db      361 DEGLIRQDRYPLRTSPQWLGPLVSDMTIAHAVALSLEAGOSTTNDPLIDVENKXTHHGNF 420
QY      354 DEGLIRQDRYPLRTSPQWLGPLVSDMTIAHAVALSLEAGOSTTNDPLIDVENKXTHHGNF 413
Db      421 QASAVNTEKTRIALALIGKLNFTQLTETMLNAGNRLPSCLAEDPSLSYHCKGIDIA 480
QY      414 QASAVNTEKTRIALALIGKLNFTQLTETMLNAGNRLPSCLAEDPSLSYHCKGIDIA 473
Db      481 AAAYTSELGHLANPYTTHVQPAEMGNQAVNSLALISARRTEANDVLSLLATHTLYCVLQ 540
QY      474 AAAYTSELGHLANPYTTHVQPAEMGNQAVNSLALISARRTESNDVLSLLATHTLYCVLQ 533
Db      541 AVDLRAMEFEFKQRPXKXXLLXQHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 600
QY      534 AIDLRALIEFEFKQRPXKXXLLXQHPGXXTXKXXKXXKXXKXRLBQTNSTYDL 593
Db      601 EPRMHDAPFSGATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRVXXKFWXAPSS 660
QY      594 EPRMHDAPFSGATGTVEKLSXXXXXVSLAAVNAKVAKAEKALSTRVXXKFWXAPSS 650
Db      661 SSPALXYLSPTRTVLYSPVREELGVKARGDVFLEKQVTTGNSVSRLEYAKSGRLNXY 720
QY      651 SSPALXYLSPTRTVLYSPVREELGVKARGDVFLEKQVTTGNSVSRLEYAKSGRLNXY 710
Db      721 LVKMLA 726
QY      711 LVKMLA 716
XX
RESULT 4
ABB07693
ID ABB07693 standard; protein; 716 AA.

```

XX ABB07693;
AC
XX 07-AUG-2003 (revised)
DT
DT 10-JUN-2002 (first entry)
XX
XX R. toruloides PAL polypeptide.
DE
XX PAL; Yeast; phenylalanine ammonia lyase; cinnamic acid; enzyme;
KW cytosolic; anti-HIV; virucide; nootropic; dermatological; gene therapy.
XX
OS Rhodosporidium toruloides.
XX
XX WC020208402-A2.
XX
XX 31-JAN-2002.
XX
XX 24-JUL-2001; 2001MO-US023270.
XX
XX 24-JUL-2000; 2000US-00624693.
XX
XX (PCBU-) PCBU SERVICES INC.
XX
XX Yoshida RK, Kootstra AB;
PI
XX WPI; 2002-268973/31.
DR
DR N-PSDB; ABA95243.
XX
XX Phenylalanine ammonia lyase polypeptide and polynucleotide useful for
PT treating mammal having disease or disorder from phenylketonuria, cancer,
PT human immunodeficiency virus infection and human cytomegalovirus
PT infection.
XX
XX Example 3; Fig 1; 135pp; English.
XX
XX The invention relates to an isolated and purified yeast phenylalanine
CC ammonia-lyase (PAL) polypeptide. The PAL encoding polynucleotide or a
CC host cell comprising the polynucleotide is useful for the production of L
CC -PAL, by adding the polynucleotide or the host cell to a composition
CC comprising trans-cinnamic acid, or trans-cinnamate and ammonia; and for
CC improving the production of PAL, its analog or another optically active
CC unnatural amino acid having PAL-like structure. The PAL polynucleotides
CC are useful for treating a mammal having a disease, disorder or condition
CC selected from phenylketonuria, cancer, human immunodeficiency virus
CC infection and human cytomegalovirus infection. The present sequence
CC represents a R. toruloides PAL polypeptide. (Updated on 07-AUG-2003 to
CC correct OS field.)
XX
XX
SQ Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 5; Length 716;
Best Local Similarity 84.0%; Pred. No. 4,5e-300;
Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MAFSLDISATXANGXNGXHAAXXAXXXXXXXXAAGXLPPTXXXTQDIVEXXLAD 60
DB 1 MAFSLDISHSFANGVAAKQAVNGAS-----TWLVAAGSHLPPTQVQDYIEKMLAA 54
QY 61 PXTDDXELDGYSLTGDVYGAARKGRXRVVDSBEIKXIDKXSEFIRKXLXNSVYGV 120
DB 55 P-TDSTLELDGYSLNIGDVVSAARKGRPVKXDSBEIKSKIDKSEFIRSQLSNVYGV 113
QY 121 TGFSGSADRTEDAIISQKALIEHOLCGVLPXTSXSFXLGRGLENSLPLEVVRGAMTIRV 180
DB 114 TGFSGSADRTEDAIISQKALIEHOLCGVLPXTSXSFXLGRGLENSLPLEVVRGAMTIRV 173
QY 181 NSLITRHSASVRLVTEFALTNFNLNGITPIVPIRGITISASGULSPUSTYAAAITHGPD 240
DB 174 NSLITRHSASVRLVTEFALTNFNLNGITPIVPIRGITISASGULSPUSTYAAAITHGPD 233
QY 241 HVHHEGEXEKIMVAREALFGLPEVVLGPKSGELGVNGTAVASASMAATLALHDAHMLSL 300
DB 234 HVHHEGEXEKILVAREMALFNLEPVVLGPKSGELGVNGTAVASASMAATLALHDAHMLSL 293

QY 301 QALTALTYEAMYGHAGSFHPLHDYTRPHPTQIEVARNIRTLLEGXSFVAVHHEEVKVD 360
DB 294 QSLTAMTYEAMYGHAGSFHPLHDYTRPHPTQIEVARNIRTLLEGXSFVAVHHEEVKVD 353
QY 361 DEGIILRODRYPLRTSPQWLGPLVSDMLHAHVALSLLEAGQSTTDNPLIDVENKXTHGNGF 420
DB 354 DEGIILRODRYPLRTSPQWLGPLVSDMLHAHVALSLLEAGQSTTDNPLIDVENKXTHGNGF 413
QY 421 QASAYXNTMEKTRIALALIGKNTFTQLTENTNAAGNGLBSCLAEPSPSYCKKGIDIA 480
DB 414 QAAVAANTMEKTRIGLAIQIGLNFQLTENTNAAGNGLBSCLAEPSPSYCKKGIDIA 473
QY 481 AAAYTSELGHLANPTTIVQPAENGNOAVNSLALISARRTEANDVLSLLATHLYCVLQ 540
DB 474 AAAYTSELGHLANPTTIVQPAENGNOAVNSLALISARRTESNDVLSLLATHLYCVLQ 533
QY 541 AVDLRAMEPEFEKQXPXXXXLXOHGXXXXTXXXXXXELXXKXKXKXRLRQNTNSYDL 600
DB 534 AIDLRAIIEFEFKQGPALIVSLIDQHFQSGAMTGNLDELEYEKXKTLAKRLBQNTNSYDL 593
QY 601 EPRMHDASFXYATGTVEXTLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXFXWAPSS 660
DB 594 VPRMHDASFXYATGTVEXTLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXFXWAPSS 650
QY 661 SSPALXYLSPTRTVLVSFVREELGVKARQGVFLGKQEVTTIGSNVSRIVYALIKSGRINXY 720
DB 651 SSPALXYLSPTRTVLVSFVREELGVKARQGVFLGKQEVTTIGSNVSRIVYALIKSGRINXY 710
QY 721 LVKMLA 726
DB 711 LVKMLA 716

RESULT 5

AAE27941
ID AAE27941 standard; protein; 716 AA.

XX AAE27941;
AC
XX 06-AUG-2003 (revised)
DT
DT 27-DEC-2002 (first entry)
XX

DE Rhodotorula toruloides PAL protein.

XX Yeast; phenylalanine ammonia lyase; PAL; EC 4.3.1.5; phenylketonuria;
KW cancer; human immunodeficiency virus infection; HIV; gene therapy; hCMV;
KW human cytomegalovirus infection; cytosolic; virucide.
XX

OS Rhodosporidium toruloides.

XX US2002102712-A1.

XX 01-AUG-2002.

XX 24-AUG-2001; 2001US-00939408.

XX 24-JUL-2000; 2000US-00624693.

XX 24-JUL-2001; 2001MO-US023270.

XX (PCBU-) PCBU SERVICES INC.

XX Yoshida RK, Kootstra AB;

XX WPI; 2002-690616/74.

XX N-PSDB; AAD45815.

XX Novel isolated and purified Rhodotorula phenylalanine ammonia lyase
PT polypeptide, useful for treating a mammal having phenylketonuria, cancer,
PT human immunodeficiency virus or human cytomegalovirus infection.

XX Example 3; Page 43-45; 74pp; English.

CC The present invention relates to yeast (e.g. Rhodotorula) phenylalanine
 CC ammonia lyase (PAL; EC 4.3.1.5) proteins and polynucleotides encoding
 CC such proteins. PAL sequences are useful for producing L-phenylalanine.
 CC They are useful for treating mammals having diseases, disorders or
 CC conditions that would benefit from treatment with PAL proteins such as
 CC phenylketonuria, cancer, human immunodeficiency virus infection (HIV) or
 CC human cytomegalovirus (hcmv) infection. Sequences of the invention are
 CC also used in gene therapy. The present sequence is R. toruloides PAL
 CC protein. (Updated on 06-AUG-2003 to correct OS field.)

XX Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 5; Length 716;

Best Local Similarity 84.0%; Pred. No. 4.5e-300;

Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MADSLSIATSYANGXNGXHAAXXASXXXXXXXAXAGSLPTTXXTOLDIVEXXLAD 60
 DB 1 MADSLSISHSFANGVASAKQAVNGAS-----TNLAAGSHLPTTQVTDIVEXMLAA 54
 QY 61 PXTDDXXELDGYSLTLDGVVGAARKGRXRVYDSDERXKIDKVEFLRXQLNNSVYGV 120
 DB 55 P-TDSTLELDGSLNLGDIVSAARKGRPVRYKDSERKIDKVEFLRSQLSMSVYGV 113
 QY 121 TGFSGADTRTEDAISLQKALLEHQLCGVLPYSXDSFXLGRGLENLPLEVVGAMTIRV 180
 DB 114 TGFSGADTRTEDAISLQKALLEHQLCGVLPSSPDSFRLGRGLENLPLEVVGAMTIRV 173
 QY 181 NSITRGSHAVRLVYVLEALNPLNHTGTPVPLRGTTASAGDISPLSYIAAIIHGPDGSKV 240
 DB 174 NSITRGSHAVRLVLEALNPLNHTGTPVPLRGTTASAGDISPLSYIAAIIHGPDGSKV 233
 QY 241 HVHHEGEXKIMKAREALFGLPEVYLGPKEGLGVNGTAVASASMATLADAMHLSLS 300
 DB 234 HVHHEGEXKILVAREMALFNLPEVYLGPKEGLGVNGTAVASASMATLADAMHLSLS 293
 QY 301 QALTALTEAMVGHAGSEHPLADVTFRPHPTQIEVARNIRTLLEGSGPAVHHEEVAKVD 360
 DB 294 QSLTALTAVAMVGHAGSEHPLADVTFRPHPTQIEVAGNIRKLESGRPVAVHHEEVAKVD 353
 QY 361 DEGLINQDSYPLRTSPQWLGPLVSDMHNHAYLSLAGOSTTNDPLDVENKTHHGNF 420
 DB 354 DEGLINQDSYPLRTSPQWLGPLVSDMHNHAYLSLAGOSTTNDPLDVENKTHHGNF 413
 QY 421 QASAVNTEKTRLALIGKLNFTQLEMLNAGNNGPSCLAEDPSLSYHCKGLDIA 480
 DB 414 QASAVNTEKTRLALIGKLNFTQLEMLNAGNNGPSCLAEDPSLSYHCKGLDIA 473
 QY 481 AAAYTSELGHANPPTTHVQPAEMNGOAVNSIALISARTIABANDVLSLLATHLVCVLQ 540
 DB 474 AAAYTSELGHANPPTTHVQPAEMNGOAVNSIALISARTIABANDVLSLLATHLVCVLQ 533
 QY 541 AAVLRAMEEFKOPFPKXXXLLXQHPGXXXTXXXXXXELXXXKXKXRLLEQTNSTYL 600
 DB 534 AIDLRALIEFEFKOPFPKXXXLLXQHPGXXXTXXXXXXELXXXKXKXRLLEQTNSTYL 593
 QY 601 EPRMHDASXAFTGVYEXISSXXXXVSLAANAMKVAAXEAKISLTRYXXFWXAPSS 660
 DB 594 VPRMHDASFAAGTVEVLS--TSLTIAAVNAKVAASASISLRFQVETFWASAST 650
 QY 661 SSPALXYLSPRTRVLYSFVREELGVKARSGDVEFLGKQVTTGSNVSRTVEYIKSGRINXV 720
 DB 651 SSPALXYLSPRTRVLYSFVREELGVKARSGDVEFLGKQVTTGSNVSRTVEYIKSGRINXV 710
 QY 721 LVEMLA 726
 DB 711 LKMLA 716

RESULT 6
 AA20653
 ID AA20653 standard; protein; 716 AA.

AC AA20653;
 XX
 XX 01-JUL-2002 (first entry)

DE Rhodotorula glutinis phenylalanine ammonia lyase (PAL).

KW Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KW cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
 KW LCP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst.

OS Rhodotorula glutinis.

PN WO200210407-A1.

PD 07-FEB-2002.

PF 23-JAN-2001; 2001WO-US002099.

PR 27-JUL-2000; 2000US-00627216.

PA (DUPO) DU FONT DE MEMOURS & CO E I.

PI Gatenby AA, Sariastani FS, Tang X, Qi WW, Vannelli T;

DR WPI; 2002-280635/32.

XX N-PSDB; AAD33075.

PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
 PT hydroxylase.

PS Example 6; Page 92-95; 139pp; English.

CC The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the
 CC plant cytochrome p-450 and the cytochrome p-450 reductase into E. coli.
 CC Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by
 CC recombinant techniques. The recombinantly produced PHCA may be used as a
 CC monomer for production of liquid crystal polymers (LCP). LCP may be used
 CC in electronic connectors and telecommunication and aerospace
 CC applications. LCP resistance to sterilizing radiation has also enabled
 CC these materials to be used in medical devices as well as chemical, and
 CC food packaging applications. The present sequence is Rhodotorula glutinis
 CC PAL protein

XX Sequence 716 AA;

Query Match 89.8%; Score 3021; DB 5; Length 716;

Best Local Similarity 84.0%; Pred. No. 4.5e-300;

Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MADSLSIATSYANGXNGXHAAXXASXXXXXXXAXAGSLPTTXXTOLDIVEXXLAD 60
 DB 1 MADSLSISHSFANGVASAKQAVNGAS-----TNLAAGSHLPTTQVTDIVEXMLAA 54
 QY 61 PXTDDXXELDGYSLTLDGVVGAARKGRXRVYDSDERXKIDKVEFLRXQLNNSVYGV 120
 DB 55 P-TDSTLELDGSLNLGDIVSAARKGRPVRYKDSERKIDKVEFLRSQLSMSVYGV 113
 QY 121 TGFSGADTRTEDAISLQKALLEHQLCGVLPYSXDSFXLGRGLENLPLEVVGAMTIRV 180
 DB 114 TGFSGADTRTEDAISLQKALLEHQLCGVLPSSPDSFRLGRGLENLPLEVVGAMTIRV 173

QY 181 NSLTRGHSARVLYVLEALTNPLNHGTTPIVPLRGITISASGDLSPSTYIAAATGHPDSKV 240
 DB 174 NSLTRGHSARVLYVLEALTNPLNHGTTPIVPLRGITISASGDLSPSTYIAAATGHPDSKV 233
 QY 241 HVXHEGEXKIMKAREALFGLPEVVLGPKSGGLVNGTAVSASMATLALDHAMLSLS 300
 DB 234 HVXHEGEXKILYAREAMALFNLPEVVLGPKSGGLVNGTAVSASMATLALDHAMLSLS 293
 QY 301 QALTALIVEMNGHAGSFPHLDVTRPHPTQIEVARNRITLLESGFAVHHEEVKVD 360
 DB 294 QSLTAMTVEAMVGHAGSFPHLDVTRPHPTQIEVAGNIRKLLESGFAVHHEEVKVD 353
 QY 361 DEGLIRQDRYPLRTSPQWLGPIVSDMIFAHAVLSLEAGOSTTNDPLIDVENKXTHGNGF 420
 DB 354 DEGLIRQDRYPLRTSPQWLGPIVSDMIFAHAVLITEAGOSTTNDPLIDVENKXTHGNGF 413
 QY 421 QASAVXNTEKTRTALALIGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 480
 DB 414 QAAAVANTWTEKTRTGLAIOIGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 473
 QY 481 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATHLVCYLQ 540
 DB 474 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARRTESNDVLSLLATHLVCYLQ 533
 QY 541 AVDIRAMEFEFKQFPPKQFXXXLLXQHPGXXXTXXXXXELXXKVKXKLKRLQSTNSYDL 600
 DB 534 AIDLRATIEFEFKQFPGPAIVSLIDQHPGSAVTGSLNRLDELVEKVKTLAKRLQSTNSYDL 593
 QY 601 EPPMHDAFSAAGTVEVEXLSXXXVSLAANAMVAAABKALSLTXVXXFXFPXAPSS 660
 DB 594 VPRMHDAFSAAGTVEVEXLS---TSLSLAANAMVAAABKALSLTXVXXFXFPXAPSS 650
 QY 661 SSPALXYLSPTRTVLYSFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 DB 651 SSPALXYLSPTRTQIILYAFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINNV 710
 QY 721 LVKMLA 726
 DB 711 LVKMLA 716
 RESULT 7
 AAE16389
 ID AAE16389 standard; protein; 716 AA.
 AC AAE16389;
 DT 09-APR-2002 (first entry)
 DB Rhodocorula glutinis wildtype phenylalanine ammonia-lyase (PAL).
 KW Phenylalanine ammonia-lyase; PAL; enzyme; citramate; food packing; PHCA;
 KW para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
 KW liquid Crystal Polymer; LCP; electronic connector; telecommunication;
 KW aerospace application; chemical; red yeast.
 OS Rhodocorula glutinis.
 PN W020011071-A2.
 PD 15-FEB-2001.
 PF 03-AUG-2000; 2000WO-US021156.
 PR 06-AUG-1999; 99US-01477192.
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 PI Tang X, Vannelli TM, Qi WW, Sarisiani S, Gatenby AA;
 DR WPI: 2002-121549/16.
 DR N-PSDB; AAD26923.
 XX

PT Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
 PT comprises converting citramate to PHCA, glucose to PHCA by phenyl ammonia
 PT -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
 PT activity.
 PS Claim 8; Page 65-67; 75pp; English.
 XX
 CC The invention relates to a method for producing para-hydroxycinnamic acid
 CC (PHCA). The method comprises converting citramate to PHCA by converting
 CC glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
 CC route or by generating a new biocatalyst possessing enhanced tyrosine
 CC ammonia-lyase (TAL) activity. The method is useful for the biological
 CC production of PHCA which is useful as a monomer for the production of
 CC liquid Crystal Polymers (LCP), where the LCP is used in electronic
 CC connectors, telecommunications and aerospace applications. LCP is also
 CC useful in medical devices, as well as chemical, and food packing
 CC applications due to its resistance to sterilizing radiation. The present
 CC sequence is red yeast, Rhodocorula glutinis (Rhodospiridium toruloides)
 CC wildtype PAL
 CC
 XX
 SQ Sequence 716 AA;
 Query Match 89.8%; Score 3021; DB 5; Length 716;
 Best Local Similarity 84.0%; Pred. No. 4,56-300;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;
 QY 1 NAPS.LDSIATSYXNGXNGXHPAAXXAXXXXXXXXAGSKLPTTXXTQLDIYEXLAD 60
 DB 1 NAPS.LDSISHSFANGVASAKQAVNGAS-----TNLVAGSHPTTIVTQVDIVKMLAA 54
 QY 61 EPTDXXELDGYSLTGLGVVGAARKGRVYRSDDEIRXKIDKVEFLRXQLNKSVYGV 120
 DB 55 P-TDSTLELDGYSLNLSGVVSHARKGRVYRSDDEIRSKIDKVEFLRQLSMVGVT 113
 QY 121 TGFSGADRTREDAISLQKALLEHQLCGVLPTSDXDFXGLGELNSPLLEVAGMTIRV 180
 DB 114 TGFSGADRTREDAISLQKALLEHQLCGVLPSDFSGFRLGELNSPLLEVAGMTIRV 173
 QY 181 NSLTRGHSARVLYVLEALTNPLNHGTTPIVPLRGITISASGDLSPSTYIAAATGHPDSKV 240
 DB 174 NSLTRGHSARVLYVLEALTNPLNHGTTPIVPLRGITISASGDLSPSTYIAAATGHPDSKV 233
 QY 241 HVXHEGEXKIMKAREALFGLPEVVLGPKSGGLVNGTAVSASMATLALDHAMLSLS 300
 DB 234 HVXHEGEXKILYAREAMALFNLPEVVLGPKSGGLVNGTAVSASMATLALDHAMLSLS 293
 QY 301 QALTALIVEMNGHAGSFPHLDVTRPHPTQIEVARNRITLLESGFAVHHEEVKVD 360
 DB 294 QSLTAMTVEAMVGHAGSFPHLDVTRPHPTQIEVAGNIRKLLESGFAVHHEEVKVD 353
 QY 361 DEGLIRQDRYPLRTSPQWLGPIVSDMIFAHAVLSLEAGOSTTNDPLIDVENKXTHGNGF 420
 DB 354 DEGLIRQDRYPLRTSPQWLGPIVSDMIFAHAVLITEAGOSTTNDPLIDVENKXTHGNGF 413
 QY 421 QASAVXNTEKTRTALALIGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 480
 DB 414 QAAAVANTWTEKTRTGLAIOIGKLNFTQLTETMLNAGNNRGLPSCLAEDPSLSYHCKGLDIA 473
 QY 481 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATHLVCYLQ 540
 DB 474 AAAYTSELGLANPVTTHVQPAEMGNQAVNSLALISARRTESNDVLSLLATHLVCYLQ 533
 QY 541 AVDIRAMEFEFKQFPPKQFXXXLLXQHPGXXXTXXXXXELXXKVKXKLKRLQSTNSYDL 600
 DB 534 AIDLRATIEFEFKQFPGPAIVSLIDQHPGSAVTGSLNRLDELVEKVKTLAKRLQSTNSYDL 593
 QY 601 EPPMHDAFSAAGTVEVEXLSXXXVSLAANAMVAAABKALSLTXVXXFXFPXAPSS 660
 DB 594 VPRMHDAFSAAGTVEVEXLS---TSLSLAANAMVAAABKALSLTXVXXFXFPXAPSS 650
 QY 661 SSPALXYLSPTRTVLYSFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINXV 720
 DB 651 SSPALXYLSPTRTQIILYAFVREELGVKARGDVFLGKQEVTTIGSNVSRITYEAIKSGRINNV 710

DR WPI: 2003-721766/68.
 DR N-PSDB: ADE36053.
 XX Recombination of nucleic acid templates, useful e.g. for producing
 PT enzymes with altered properties, is based on cyclic extension of unpaired
 PT primers.
 XX
 PS Disclosure; SEQ ID NO 2; 481bp; English.
 XX
 CC The present invention describes a method for the recombination of nucleic
 CC acid (NA) templates. The method comprises: (a) providing at least two
 CC double-stranded templates (T1, T2) having different 5' and 3' regions in
 CC their sense strands; (b) contacting the templates with at least one each
 CC of forward and reverse primers that anneal, respectively, only to the 3'-
 CC region of the antisense strand of T1 and to the 3'-region of the sense
 CC strand of T2; (c) extending the primers by no more than 1000 nucleotides
 CC (nt); (d) separating extended primers from their templates; (e)
 CC reannealing the extended primers, to either template; and (f) repeating
 CC cycles of steps (b)-(e) until at least one full-length extension product,
 CC i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
 CC method starting from at least one each of antisense and sense single-
 CC stranded templates; (2) generating a recombined polypeptide (I) with
 CC altered properties by expressing recombination products of the new
 CC processes and screening the polypeptides formed for properties different
 CC from those of polypeptides from either of the templates; and (3) (1)
 CC produced by method (2). The methods are useful for the recombination of
 CC nucleic acid templates of interest, and generation of a recombined
 CC polypeptide having altered properties. The method can be used to make
 CC combined genes that express proteins e.g. enzymes, cytokines, growth
 CC factors, viral proteins or microbial antigens with altered properties,
 CC e.g. stability, activity or specificity. The present sequence is used in
 CC the exemplification of the present invention.
 XX
 XX Sequence 716 AA:
 SQ

Query Match 89.8%; Score 3021; DB 7; Length 716;
 Best Local Similarity 84.0%; Pred. No. 4,5e-300;
 Matches 610; Conservative 25; Mismatches 81; Indels 10; Gaps 3;

QY 1 MABSLDISATXKXNGXKXGAXXASXXXXXXKXGXXLPFTXXYTOLDIYEXXAD 60
 Db 1 MABSLDISATXKXNGXKXGAXXASXXXXXXKXGXXLPFTXXYTOLDIYEXXAD 60
 QY 61 PXTDDXXELDGLTIGDVGAARKGRXVRVXDSDEIRKXIDKSYEFRLXOLXNSVYGV 120
 Db 55 P-TDSTLELDGSLNLDVVSARKGRPVVRKDSDEIRKIDKSYEFRLXOLXNSVYGV 113
 QY 121 TGFSGADRTTEPAISLOKALEHOLCGVLPXSXSFKXGRLENSLLEVVRRGAMTRV 180
 Db 114 TGFSGADRTTEPAISLOKALEHOLCGVLPXSXSFKXGRLENSLLEVVRRGAMTRV 173
 QY 181 NSLTRGSAVRVLYVLEALTNFLNHTPIVPLRGRTISASGDLSPSYIAAIIHGPDSKV 240
 Db 174 NSLTRGSAVRVLYVLEALTNFLNHTPIVPLRGRTISASGDLSPSYIAAIIHGPDSKV 233
 QY 241 HVHHEGEEKIMKARELIAFGLEPVYLGKXEGELGVNGTAVASASMATALDAMLSLS 300
 Db 234 HVHHEGEEKILVAREMALFNEPVYLGKXEGELGVNGTAVASASMATALDAMLSLS 293
 QY 301 QALTALTVAMVGHAGSFHPLHDVTRPHPTQIEVARNIPTLLESGXFVHHEEVAKD 360
 Db 294 QSLTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVARNIPTLLESGXFVHHEEVAKD 353
 QY 361 DEGLIQDRYPLRTSQNGPLVSDMITAHAVLSLBAQSTTNDPLIVENKXTHHGNGF 420
 Db 354 DEGLIQDRYPLRTSQNGPLVSDMITAHAVLSLBAQSTTNDPLIVENKXTHHGNGF 413
 QY 421 QASAVXNTEKTRIALALIGKLNFTQLEMLNAGNNGRGLPCLAAEDPSLSYHCKGDIA 480
 Db 414 QAAAVANTEKTRILGALIGKLNFTQLEMLNAGNNGRGLPCLAAEDPSLSYHCKGDIA 473
 QY 481 AAAYSELGLHANPVTTHVQPAEMGNQAVNSIALISARTRAEANDVLSLLIATHLVCLQ 540

Db 474 AAAYSELGLHANPVTTHVQPAEMGNQAVNSIALISARTRAEANDVLSLLIATHLVCLQ 533
 QY 541 AVDLRAMEFEFKQFXKXKXLLXOHFGXXXTXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
 Db 534 AILRLAIEFEFKQFQFPAVSLIDQHFSGAMGSLRDELVEKNKTLAKRLKLEQNSYDL 593
 QY 601 EPRWDAFSGXGTGTVEKLSXXXXXVSLAANAMKVAAEKAISLTXXVXXFXKXSS 660
 Db 594 VPRWDAFSGXGTGTVEKLS--TSLAANAMKVAAEKAISLTXXVXXFXKXSS 650
 QY 661 SSPALXYSPPRTVLSFVREELGVKARQGVFLGKQETIGSNVSRITYEAKSGRIYXV 720
 Db 651 SSPALXYSPPRTVLSFVREELGVKARQGVFLGKQETIGSNVSKITYEAKSGRIYXV 710
 QY 721 LVXKLA 726
 Db 711 LKXKLA 716
 RESULT 10
 ID AAE20668 standard; protein, 716 AA.
 XX
 AC AAE20668;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE R. glutinis phenylalanine ammonia lyase (PAL) mutant, Epl8km-6.
 XX
 KW Para-hydroxycinnamic acid; PHCA; food packaging; electronic connector;
 KW phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
 KW cytochrome P-450; cytochrome P-450 reductase; liquid crystal polymer;
 KW ICP; telecommunication; medical device; aerospace application; enzyme;
 KW biocatalyst; mutant; variant; mutcin.
 XX
 XX Rhodotorula glutinis.
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 540
 FT /note= "Wild-type Ile substituted with Thr"
 XX
 EN WO200210407-A1.
 XX
 PD 07-FEB-2002.
 XX
 PF 23-JAN-2001; 2001WO-US002099.
 PR 27-JUL-2000; 2000US-00627216.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PI Gatenby AA, Sariastani FS, Tang X, Qi WW, Vannelli T;
 XX WPI, 2002-280635/32.
 DR N-PSDB: AAD33100.
 XX
 PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
 PT polypeptide, or mutant TAL polypeptide, useful for producing
 PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
 PT hydroxylase.
 XX
 PS Disclosure; Page; 139pp; English.
 XX
 CC The invention relates to methods for biological production of para-
 CC hydroxycinnamic acid (PHCA). The invention relates to developing of a new
 CC biocatalyst for conversion of glucose to PHCA by incorporation of the
 CC wild type phenylalanine ammonia lyase (PAL; EC 4.3.1.5) from the yeast
 CC Rhodotorula glutinis into Escherichia coli underlying the ability of the
 CC wildtype PAL to convert tyrosine to PHCA. The invention is also directed
 CC to developing a new biocatalyst for conversion of glucose to PHCA by
 CC incorporation of the wildtype PAL which possesses enhanced tyrosine
 CC ammonia lyase (TAL) activity from the yeast Rhodotorula glutinis plus the

CC plant cytochrome P-450 and the cytochrome P-450:reductase into *E. coli*.
CC Nucleic acid encoding mutant P4L enzyme is useful for producing PHCA by
CC recombinant techniques. The recombinantly produced PHCA may be used as a
CC monomer for production of liquid crystal polymers (LCP). LCP may be used
CC in electronic connectors and telecommunication and aerospace
CC applications. LCP resistance to sterilizing radiation has also enabled
CC these materials to be used in medical devices as well as chemical, and
CC food packaging applications. The present sequence is Rhodococcus glutinis
CC mutant P4L. Note: This sequence is not shown in the specification,
CC however it is constructed based on the P4L enzyme shown as SEQ.ID.NO:8
CC (AAE20633) in the sequence listing

50 Sequence 716 AA;

Query Match	89.7%;	Score 3019;	DB 5;	Length 716;
Best Local Similarity	84.0%;	Pred. No. 7,2e-300;		
Matches 610; Conservative	24;	Mismatches 82;	Indels 10;	Gaps 3;

Qy	1	MASLDISITSSXANGXNKGXHAAXXASXXXXXXXXXXXXXGSLPTTXXLTQDLIVEXXAD	6
Db	1	MASLDISITSSXANGXNKGXHAAXXASXXXXXXXXXXXXXGSLPTTXXLTQDLIVEXXAD	54
Qy	61	PXDDXELLEDGSLITIGDYVGAARKRXXVXDSEIRKIDKSVEPLRXOLXNSYGYT	12
Qy	55	P-TDSTLELDGSLINDGVSAARKRPRVXDSEIRKIDKSVEPLRXOLXNSYGYT	11
Db	121	TGFGSADRTEDAILOKALLEHQCVLPYXSDFXGRLGJMSLPLEYVRGAMTTRV	18
Qy	114	TGFGSADRTEDAILOKALLEHQCVLPBSFDSFRLGRLGJMSLPLEYVRGAMTTRV	17
Db	181	NSITRGHSAVRLVLEBALNFINHGITPVPVRGTTASGDLSPSYAAAITGHPDSKV	24
Qy	174	NSITRGHSAVRLVLEBALNFINHGITPVPVRGTTASGDLSPSYAAAITGHPDSKV	23
Qy	241	HYHHEKEXKIMKREAIATFGLPEVPLVGFKEGGLVNGAVASASMAATLALHDAHMLSLIS	30
Db	234	HYHHEKEXKILYAREMALFNLEPVLGKREGGLVNGAVASASMAATLALHDAHMLSLIS	29
Qy	301	QALITALTVEANGHAGSFHPLHDVTRPHPTQIEVARNITLLBGSXPAVHHEEVKYXD	36
Db	294	QSTANTATVAMGHAGSFHPLHDVTRPHPTQIEVAGNITRLLBGSXPAVHHEEVKYXD	35
Qy	361	DEGLMDEEPLRTSPQWGLPLVDMIHAAVLSLEAGOSTDNDPLIVENKXTHGNGF	42
Db	354	DEGLMDEEPLRTSPQWGLPLVSDIHHAAVLTLEAGOSTDNDPLIVENKXTHGNGF	41
Qy	421	QASAVXNTEKTRLATALIGKLNFTQLTMLNAGNMGJPSCLAEDPSLSYHCKGLDIA	48
Db	414	QAAAVVNTMEKTRLGAIOIKLNFTQLTMLNAGNMGJPSCLAEDPSLSYHCKGLDIA	47
Qy	481	AAAYTSELHLNPPVTHOPAPAMGQANSLALISARTAAANVLSILLATHTHCYLO	54
Db	474	AAAYTSELHLNPPVTHOPAPAMANOANSLALISARTTESNVLSILLATHTHCYLO	53
Qy	541	AYDLRAMEEFKKQFXPXXXXLLXOHFGXXXTXXXXXLXXKXVKXLAKRLEQNSYDL	60
Db	534	AIDLRTATEEFKKQFPAIVSLIDHFGSAMGSMRLDELVEKXKTKLAKRLEQNSYDL	59
Qy	601	EEPMHAAFSKACTVEXLSSXXXXXVSLAAVANKVLAABEALISLTXXYRXKFXAASS	66
Db	594	VPMHDAFSTAACTIVEVLSS--TSLSLAANAMVLAABEALISLTQOVRETFSAAST	65
Qy	661	SSPALXYLSBRTFVLSFVREELGVARAGDVFLLKQEVTTISANVSRIYEAIKSGRINXV	72
Db	651	SSPALXYLSBRTQILYAFVREELGVARAGDVFLLKQEVTTISANVSRIYEAIKSGRINXV	71
Qy	721	LVKMLA 726	
Db	711	LVKMLA 716	

RESULT 11
AAE20654

ID AAE20654 standard; protein; 716 AA.

AC AAE20654

DT 01-JUL-2002 (first entry)

DE Rhodotorula glutinis phenylalanine ammonia lyase (PAL) mutant.

KM Para-hydroxychromamic acid; PHCA food packaging; electronic connector;
KM phenylalanine ammonia lyase; PAL; glucose; tyrosine ammonia lyase; TAL;
KM cytochrome p-450; cytochrome p-450 reductase; liquid crystal polymer;
KM LCP; telecommunication; medical device; aerospace application; enzyme;
KM biocatalyst; mutant; variant; muten.

Rhodotorula glutinis.

Synthetic

PN WO200210407-A1

PD 07-FEB-2002.

PF 23-JAN-2001; 2001WO-US002099

PR 27-JUL-2000; 2000US-00627216

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Gatendy AA, Sariastani FS, Tang X, Qi WW, Vannelli T, XY

DR WPI; 2002-280635/32.
DB N-PENDB; AAD33076

Novel multi agent

PT Novel nucleic acid encoding truncated mutant tyrosine ammonia lyase (TAL)
PT polypeptide, or mutant TAL polypeptide, useful for producing
PT parahydroxycinnamic acid in recombinant host cell lacking cinnamate
PT hydroxylase.

PS Claim 2; Page 98-101; 139pp; English.

The invention relates to methods for biological production of para-hydroxyphenylacetic acid (PHCA). The invention relates to developing of a new biocatalyst for conversion of glucose to PHCA by incorporation of the wild type phenylalanine ammonia lyase (PAL, EC 4.3.1.5) from the yeast *Rhodotorula glutinis* into *Escherichia coli* underlying the ability of the wildtype PAL to convert tyrosine to PHCA. The invention is also directed to developing a new biocatalyst for conversion of glucose to PHCA by incorporation of the wildtype PAL which possesses enhanced tyrosine ammonia lyase (TAL) activity from the yeast *Rhodotorula glutinis* plus the plant cytochrome p-450 and the cytochrome p-450 reductase into *E. coli*. Nucleic acid encoding mutant TAL enzyme is useful for producing PHCA by recombinant techniques. The recombinantly produced PHCA may be used as monomer for production of liquid crystal polymers (LCP). LCP may be used in electronic connectors and telecommunication and aerospace applications. LCP resistance to sterilising radiation has also enabled these materials to be used in medical devices as well as chemical, and food packaging applications. The present sequence is *Rhodotorula glutinis* PAL mutant

SQ Sequence 716 AA;

Query Match 89.7%; Score 3019; DB 5; Length 716;
Best Local Similarity 84.0%; Pred. No. 7.2e-300;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3,

[illegible]

```

Db      114 TGFSGSADTRTEDAISTQKALLHQLGCVLPSSFDFRLGRGJENSLPEVVRGAMTIRV 173
QY      181 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 240
Db      174 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 233
QY      241 HVHEGXEKIMXAREALFGLPEVYLGPKEGGLVNGTAVASMTALTALHDAHMLSLLS 300
Db      234 HVHEGXEKILYAREALFGLPEVYLGPKEGGLVNGTAVASMTALTALHDAHMLSLLS 293
QY      301 QALTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGGSPFAVHHEBEVKYD 360
Db      294 QSLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGGSPFAVHHEBEVKYD 353
QY      361 DEGLIRQDRYPLRTSPQWIGPLVSDMIHAHVAVLSIAGOSTDNPILIDVENKXTHGGNF 420
Db      354 DEGLIRQDRYPLRTSPQWIGPLVSDMIHAHVAVLSIAGOSTDNPILIDVENKXTHGGNF 413
QY      421 QASAVNTMEKTRILALIGKLNFTQLEMLNAGNRGLPSCIAEDPGLSYHCKGLDIA 480
Db      414 QAAVANTMEKTRILALIGKLNFTQLEMLNAGNRGLPSCIAEDPGLSYHCKGLDIA 473
QY      481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARTAEANDVLSLLATHLVCYLQ 540
Db      474 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARTAESNDVLSLLATHLVCYLQ 533
QY      541 AVDLRAMEFEFKQFPKXKXLLXHPFGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600
Db      534 AIDRATGTEFEFKQFPKXKXLLXHPFGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 593
QY      601 EPRWHDASXATGTVVEVXLSKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 660
Db      594 VPRWHDASXATGTVVEVXLSKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 650
QY      661 SSPALXLSPTRTLYLFFVREBELGYKARQGVFLGQVETLGSNNRYIAEATKSRINXV 720
Db      651 SSPALXLSPTRTLYLFFVREBELGYKARQGVFLGQVETLGSNNRYIAEATKSRINXV 710
QY      721 LVKMLA 726
Db      711 LKMLA 716

RESULT 12
AAE16390
ID AAE16390 standard; protein; 716 AA.
AC AAE16390;
XX
DT 09-APR-2002 (first entry)
XX
DE Rhodotorula glutinis mutant phenylalanine ammonia-lyase (PAL).
XX
KW Phenylalanine ammonia-lyase; PAL; enzyme; cinnamate; food packing; PHCA;
KW para-hydroxycinnamic acid; tyrosine ammonia-lyase; TAL; medical device;
KW Liquid Crystal Polymer; LCP; electronic connector; telecommunication;
KW aerospace application; chemical; red yeast; mutant; mutcin.
XX
OS Rhodotorula glutinis.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 540 /note= "Wild type Ile substituted with Thr"
XX
PN WO20011071-A2.
XX
PD 15-FEB-2001.
XX
PF 03-AUG-2000; 2000MO-US021156.
XX
PR 06-AUG-1999; 99US-0147719P.

```

```

XX (DUPO) DU POINT DE NEMOURS & CO E. I.
XX Tang X, Vannelli TM, Qi WW, Sariafiani S, Gatenby AA;
XX WPI; 2002-121549/16.
XX N-PSDB; AAD26924.
XX
PT Producing para-hydroxycinnamic acid for producing liquid crystal polymer,
PT comprises converting cinnamate to PHCA, glucose to PHCA by phenyl ammonia-
PT -lyase route, or generating a biocatalyst with tyrosine ammonia-lyase
PT activity.
XX
PS Claim 14; Page 68-70; 75pp; English.
XX
CC The invention relates to a method for producing para-hydroxycinnamic acid
CC (PHCA). The method comprises converting cinnamate to PHCA by converting
CC glucose to phenylalanine to PHCA through the phenyl ammonia-lyase (PAL)
CC route or by generating a new biocatalyst possessing enhanced tyrosine
CC ammonia-lyase (TAL) activity. The method is useful for the biological
CC production of PHCA which is useful as a monomer for the production of
CC liquid crystal polymers (LCP), where the LCP is used in electronic
CC connectors, telecommunications and aerospace applications. LCP is also
CC useful in medical devices, as well as chemical, and food packing
CC applications due to its resistance to sterilizing radiation. The present
CC sequence is red yeast, Rhodotorula glutinis (Rhodospiridium toruloides)
CC mutant PAL derived by substituting Ile at position 540 with Thr. Mutant
CC PAL enzyme possesses enhanced TAL activity
XX
SQ Sequence 716 AA:
XX
Query Match 89.7%; Score 3019; DB 5; Length 716;
Best Local Similarity 84.0%; Pred. No. 7.2e-300;
Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;
QY 1 MAPSLDSTISXANXKXNXXHAXXAXSXXXXXXXXXXAAGSLTXXXTQTDIYEXXAD 60
Db 1 MAPSLDSTISHSFANGVASKQAVNGAS-----TNLAAGSHLTQYTDIYEXKLLA 54
QY 61 PXTDDXELDGYSLTIGDVGAARKGRXRYXDSDEIRKXIDKSVFRLQKXNSYGVY 120
Db 55 P-TDSTLELDGYSLNIGDVGAARKGRPVRYVDSDEIRSKIDKSVFRLQKXNSYGVY 113
QY 121 TGRGGSADTRTEDAISTQKALLHQLGCVLPSSDFRLGRGJENSLPEVVRGAMTIRV 180
Db 114 TGRGGSADTRTEDAISTQKALLHQLGCVLPSSDFRLGRGJENSLPEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 240
Db 174 NSLTRGSAVRLVLEALTNFLNHGITPIVPLRGITISAGDLSPLSYIAAATGHPDSKV 233
QY 241 HVHEGXEKIMXAREALFGLPEVYLGPKEGGLVNGTAVASMTALTALHDAHMLSLLS 300
Db 234 HVHEGXEKILYAREALFGLPEVYLGPKEGGLVNGTAVASMTALTALHDAHMLSLLS 293
QY 301 QALTALTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGGSPFAVHHEBEVKYD 360
Db 294 QSLTAMTVEAMVGHAGSFHPLHDVTRPHPTQIEVARIRITLGGSPFAVHHEBEVKYD 353
QY 361 DEGLIRQDRYPLRTSPQWIGPLVSDMIHAHVAVLSIAGOSTDNPILIDVENKXTHGGNF 420
Db 354 DEGLIRQDRYPLRTSPQWIGPLVSDMIHAHVAVLSIAGOSTDNPILIDVENKXTHGGNF 413
QY 421 QASAVNTMEKTRILALIGKLNFTQLEMLNAGNRGLPSCIAEDPGLSYHCKGLDIA 480
Db 414 QAAVANTMEKTRILALIGKLNFTQLEMLNAGNRGLPSCIAEDPGLSYHCKGLDIA 473
QY 481 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARTAEANDVLSLLATHLVCYLQ 540
Db 474 AAAYTSELGLANPVTTHVQPAEMGQAVNSLALISARTAESNDVLSLLATHLVCYLQ 533
QY 541 AVDLRAMEFEFKQFPKXKXLLXHPFGKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 600

```

Db 534 AILRLATFEFFKQFGPAIVSLIDHFGSAMTGSNLRDELVEKUNKTLAKLEQTNVDL 593
QY 601 EPRWHDASFATGTVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXFXWAPSS 660
Db 594 VPRWHDASFATGTVVEXLS--TSLSIAAVNAKVAAXEKAISLTRVXETFWMSAAT 650
QY 661 SSPALXYLSPPRTVYVSFFREEGVKARPGDVFELKQEVTTGSNVRILEYAIKSGRINXV 720
Db 651 SSPALXYLSPPRTQILYAFVREEGVKARPGDVFELKQEVTTGSNVRILEYAIKSGRINNV 710
QY 721 LVKMLA 726
Db 711 LKMLA 716
RESULT 13
ABG71955
ID ABG71955 standard; protein; 716 AA.
XX
AC ABG71955;
XX
DT 27-JAN-2003 (first entry)
XX
De R. glutinis PAL/TAL naturally occurring variant.
XX
KM Enzyme; phenylalanine ammonia-lyase; PAL; phenylalanine hydroxylase; PAH;
KM para-hydroxycinnamic acid; tyrosine ammonium lyase; TAL; PAL/TAL;
XX tyrosine; cinnamic acid; PHCA; liquid crystal polymer.
XX
OS Rhodornula glutinis.
XX
PN WO200290523-A2.
XX
PD 14-NOV-2002.
XX
PF 03-MAY-2002; 2002MO-US018551.
XX
PR 04-MAY-2001; 2001US-0288701P.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Qi WW, Sariastani FS, Tang X;
XX
XX WPI; 2003-058928/05.
DR N-PSDB; ABS56543.
XX
PT Novel recombinant host for producing para-hydroxycinnamic acid and
PT tyrosine, comprises gene encoding tyrosine ammonium lyase activity and
PT gene encoding phenylalanine hydroxylase activity.
XX
PS Disclosure; Page 67-69; 69pp; English.
XX
XX The invention relates to a recombinant host comprising at least one gene
CC encoding a tyrosine ammonium lyase (TAL) activity, and at least one gene
CC encoding a phenylalanine hydroxylase (PAH) activity. Also included are:
CC (1) a method for producing para-hydroxycinnamic acid involving: (1)
CC providing the recombinant host; and (2) growing the recombinant organism
CC in the presence of a fermentable carbon substrate and thus producing para
CC -hydroxycinnamic acid; (2) method for producing tyrosine involving: (1)
CC providing a recombinant organism comprising at least one gene encoding a
CC phenylalanine hydroxylase activity; and (2) growing the recombinant
CC organism in the presence of a fermentable carbon substrate to produce
CC tyrosine. The recombinant host cell is used for producing para-
CC hydroxycinnamic acid and tyrosine. Increasing the carbon flow into the
CC production of para-hydroxycinnamic acid (PHCA) which is useful as a
CC monomer for the production of liquid crystal polymer. The present
CC sequence is a naturally occurring variant of the R. glutinis
CC Phenylalanine ammonia-lyase, PAL, which has TAL activity (PAL/TAL)
XX
XX Sequence 716 AA;
XX
Query Match 89.7%; Score 3019; DB 6; Length 716;
Best Local Similarity 84.0%; Pred. No. 7,2e-300;

Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;
QY 1 MAPSLDISATSGXANGXNKGHAAXASXXXXXXAAAGSXLPTTXXTQDIDVEXXLD 60
Db 1 MAPSLDISATSHSPANGVAVAKQAVNGAS-----TTLAVAGSHLPTTQVTDIDVEXMLAA 54
QY 61 PXTDXXELIDGSLTLGDVVGAAKRGKRVYXDSDEITKXDKSVEFLRXDLXNSVYGV 120
Db 55 P-TDSTLELDGSLNLGVVSAARRGPRVYKDSDEIRSKIDKVEFLRSQLSMSVYGV 113
QY 121 TGFSGSADRTEDATSLQKALIEHQCGVLPFSXDSFYLGRGLNSLPLEVVRGAMTIRV 180
Db 114 TGFSGSADRTEDATSLQKALIEHQCGVLPFSDFSLGRGLNSLPLEVVRGAMTIRV 173
QY 181 NSLTRGSAVRLVLTBALTNFLNHSITPIVPLRGITASAGDLSPLSTIAAITGHPDSKV 240
Db 174 NSLTRGSAVRLVLTBALTNFLNHSITPIVPLRGITASAGDLSPLSTIAAITGHPDSKV 233
QY 241 HYHHEGEXEKIMKARFAIALFGLPEVVLGPRKEGLGVNGTAVASAMATLALHDAMLSLS 300
Db 234 HYHHEGEXEKILYAFENMALFNLEPVLGPRKEGLGVNGTAVASAMATLALHDAMLSLS 293
QY 301 QALTALTYEAVYGAHGSFHPFLHDVTRBHPQIEVARNITLLEGSXFAYVHHEEYKVD 360
Db 294 QSLTAMTYEAVYGAHGSFHPFLHDVTRBHPQIEVAGNIRKLLGSRFAVHHEEYKVD 353
QY 361 DEGLIRODRYPRLRTSPQWGLVSMITHAAVLSLEAGQSTNDPLIDVENKTHHGNF 420
Db 354 DEGLIRQDYPRLRTSPQWGLVSDLIHAAVLTIEAGQSTNDPLIDVENKTHHGNF 413
QY 421 QASAVXNTEKTRIALALIGKLNFTQLEMLNAGNRLPSCLAAEDPSLSYHCKGDLA 480
Db 414 QAAAVANTMEKTRIGLAQIGKLNFTQLEMLNAGNRLPSCLAAEDPSLSYHCKGDLA 473
QY 481 AAAYTSELGHLANPTTHVOPAEKNOAVNSLALISARTBEANDVLSLIATILYCYLQ 540
Db 474 AAAYTSELGHLANPTTHVOPAEKNOAVNSLALISARTBESNDVLSLIATILYCYLQ 533
QY 541 AYDLRAMEFEFKQXPXXXXLXQHPGXXXXXXKXXKXXLXRLBEOTNSYDL 600
Db 534 AILDRAATEFFKQFGPAIVSLIDHFGSAMTGSNLRDELVEKUNKTLAKLEQTNVDL 593
QY 601 EPRWHDASFATGTVVEXLSXXXXXVSLAAVNAKVAAXEKAISLTRVXXFXWAPSS 660
Db 594 VPRWHDASFATGTVVEXLS--TSLSIAAVNAKVAAXEKAISLTRVXETFWMSAAT 650
QY 661 SSPALXYLSPPRTVYVSFFREEGVKARPGDVFELKQEVTTGSNVRILEYAIKSGRINXV 720
Db 651 SSPALXYLSPPRTQILYAFVREEGVKARPGDVFELKQEVTTGSNVRILEYAIKSGRINNV 710
QY 721 LVKMLA 726
Db 711 LKMLA 716
RESULT 14
ADE36056
ID ADE36056 standard; protein; 716 AA.
XX
AC ADE36056;
XX
DT 29-JAN-2004 (first entry)
XX
XX Rhodosporidium glutinis Epi8Km-6 mutant PAL protein SEQ ID NO.4.
XX
XX recombination; nucleic acid template; recombinant polypeptide;
XX altered property; combined gene.
XX
OS Rhodosporidium.
XX
XX WO2003072743-A2.
PN
XX
XX 04-SEP-2003.

XX 26-FEB-2003; 2003MO-US005708.
 XX 26-FEB-2002; 2002US-0360279P.
 XX (UPO) DU PONT DE NEMOURS & CO E. I.
 PA Milano J, Tang X;
 PI WPI; 2003-721766/68.
 DR N-PSDB; ADE36055.
 XX Recombination of nucleic acid templates, useful e.g. for producing
 PT enzymes with altered properties, is based on cyclic extension of unpaired
 PT primers.
 PS Disclosure; SEQ ID NO 4; 481bp; English.
 XX The present invention describes a method for the recombination of nucleic
 CC acid (NA) templates. The method comprises: (a) providing at least two
 CC double-stranded templates (T1, T2) having different 5' and 3' regions in
 CC their sense strands; (b) contacting the templates with at least one each
 CC of forward and reverse primers that anneal, respectively, only to the 3'-
 CC region of the antisense strand of T1 and to the 3'-region of the sense
 CC strand of T2; (c) extending the primers by no more than 1000 nucleotides
 CC (nt); (d) separating the extended primers from their templates; (e)
 CC reannealing the extended primers, to either template; and (f) repeating
 CC cycles of steps (b)-(e) until at least one full-length extension product,
 CC i.e. a recombination of T1 and T2, is obtained. Also described: (1) a
 CC method starting from at least one each of antisense and sense single-
 CC stranded templates; (2) generating a recombinant polypeptide (1) with
 CC altered properties by expressing recombination products of the new
 CC processes and screening the polypeptides formed for properties different
 CC from those of polypeptides from either of the templates; and (3) (1)
 CC produced by method (2). The methods are useful for the recombination of
 CC nucleic acid templates of interest, and generation of a recombinant
 CC polypeptide having altered properties. The method can be used to make
 CC combined genes that express proteins e.g. enzymes, cytokines, growth
 CC factors, viral proteins or microbial antigens with altered properties,
 CC e.g. stability, activity or specificity. The present sequence is used in
 CC the exemplification of the present invention.
 XX
 XX Sequence 716 AA;

Query Match 89.7%; Score 3019; DB 7; Length 716;
 Best Local Similarity 84.0%; Pred. No. 7.2e-300;
 Matches 610; Conservative 24; Mismatches 82; Indels 10; Gaps 3;

QY 1 MABSLDSIATSGXNGXNGXHAAXSXXXXXXXXXXAXXGSLPTTXXTOLDIYEXXLLAD 60
 DB 1 MABSLDSIATSGXNGXNGXHAAXSXXXXXXXXXXAXXGSLPTTXXTOLDIYEXXLLAA 54
 QY 61 PTDDXXELDGYSLTIGDVGAARKGRXVRVXDSDEIRAKIDKSYEPFLRXQXNVYCVT 120
 DB 55 P-IDSTLELDGYSLNIGDVVSAARKGRPVXVDSDEIRAKIDKSYEPFLRXQXNVYCVT 113
 QY 121 TGGGASADTDEDAISIQKALLFHQICVLP7TSDXSFKIGRLGSLNLEPLEYVRGAMTRV 180
 DB 114 TGGGASADTDEDAISIQKALLFHQICVLP7TSDXSFKIGRLGSLNLEPLEYVRGAMTRV 173
 QY 181 NSLTRGSAVRLVLEALTFNLHGITPIVPLRGITISAGDSPISTYIAAATIGHPDSKV 240
 DB 174 NSLTRGSAVRLVLEALTFNLHGITPIVPLRGITISAGDSPISTYIAAATIGHPDSKV 233
 QY 241 HVXHEGKEMKARREALFGLPEVVLGPEXGIGLVNGTAVASAMATLALDHAKLILLS 300
 DB 234 HVXHEGKEMKARREALFGLPEVVLGPEXGIGLVNGTAVASAMATLALDHAKLILLS 293
 QY 301 QALTLALTEAMVGHAGSFHPLDVTFRPHPTQIEVARNIRTLGSGXFAVHEEPEVKYD 360
 DB 294 QSLTAMTEAMVGHAGSFHPLDVTFRPHPTQIEVARNIRTLGSGXFAVHEEPEVKYD 353
 QY 361 DEGLIRDRYPLRTSPQWGLPLVSDMIHAHVALISLHAGSGSTTDNPLIDVENKXTHHGNF 420

DB 354 DEGLIRDRYPLRTSPQWGLPLVSDMIHAHVALISLHAGSGSTTDNPLIDVENKXTHHGNF 413
 QY 421 QASAVXNTWEKTRALALIGLNTOLTETMINAGMNGLSCLAAEPSPSTHYCKGLDIA 480
 DB 414 QAAAVANTWEKTRALALIGLNTOLTETMINAGMNGLSCLAAEPSPSTHYCKGLDIA 473
 QY 481 AAAYTSELGLANPTTIVQPAEMGNOAVNSIALISARPTAEANDVSLTLATLYCVLQ 540
 DB 474 AAAYTSELGLANPTTIVQPAEMGNOAVNSIALISARPTAEANDVSLTLATLYCVLQ 533
 QY 541 AVDLRAMEFEFKQCPYXXXXLLXQHPGXYYTXXXXXXELXXKXXKXLRIGQNSYDL 600
 DB 534 AIDLRAFEFEFKQCPYXXXXLLXQHPGXYYTXXXXXXELXXKXXKXLRIGQNSYDL 593
 QY 601 EPRMHDASFAGTVEVLSXXXXXXVSLAAVNAWTVAXAEKASITLRXVXXFWXAPSS 660
 DB 594 VPRMHDASFAGTVEVLSXXXXXXVSLAAVNAWTVAXAEKASITLRXVXXFWXAPSS 650
 QY 661 SSPALXYLSPRTVLYSFVREELGVKARQGVFLGKQEVITIGSNVSRIVYAIKSGRINXV 720
 DB 651 SSPALXYLSPRTVLYSFVREELGVKARQGVFLGKQEVITIGSNVSRIVYAIKSGRINXV 710
 QY 721 LYXMLA 726
 DB 711 LYXMLA 716

RESULT 15

AA81099
 ID AAP81099 standard; protein; 716 AA.

XX AAP81099;
 AC

DT 12-DEC-1990 (first entry)
 XX

DE Sequence of Rhodospiridium toruloides IFO 559 cells L-Phenylalanine
 DE ammonia lyase (PAL).

XX Aspartame; artificial sweetener; enzyme.
 XX

OS Rhodospiridium toruloides.
 XX

PN BP260919-A.
 XX

PD 23-MAR-1988.
 XX

PF 15-SEP-1987; 87EP-00308119.
 XX

PR 16-SEP-1986; 86JP-00215864.
 XX

PA (MITR) MITSUI TOKAI CHEM INC.
 XX

PI Fukuhara N, Yoshino S, Yamamoto K, Se T, Sone S, Nakajima Y;
 PI Suzuki M, Makiguchi N;
 XX

DR WPI; 1988-079223/12.
 DR N-PSDB; AA81099.

PT New L-phenylalanine ammonia lyase and encoding DNA sequence - useful for
 PT producing L-phenylalanine from cinnamic acid and ammonia.
 XX

PS Claim 1; Page 21-23; 37p; English.
 XX

CC PAL is useful in prodn. of L-phenylalanine from cinnamic acid and an NH3
 CC donor (process claimed). L-phe is an essential AA and starting material
 CC for sweeter aspartame. E.coli transformants will produce PAL without
 CC induction by expensive AAs, no surfactant is needed to increase cell wall
 CC permeability
 XX

XX Sequence 716 AA;
 SQ

Query Match 89.7%; Score 3018; DB 1; Length 716;

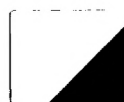
Best Local Similarity 83.9%; Pred. No. 9,1e-300;
Matches 609; Conservative 26; Mismatches 81; Indels 10; Gaps 3;

```

QY 1 MAPSLDSIATSYANGXNGXHAAXXAXXXXXXXXAGSLPTTXXTQLDIVEXXAD 60
Db 1 MAPSLDSISHSPANGVASAKAVNGAS-----TNLAAGSHLPTTQVTDIVEXKLLA 54
QY 61 EYTDXXELDGLSLGLGVYGAARKGRVYXDSDEIRXKIDKSEFLRXQLXSVYGV 120
Db 55 P-TDSTLELDGSLNLDGVVSAARKGRVYXDSDEIRSKIDKSEFLRQLSMVYGV 113
QY 121 TGGGSGADRTEDATSLQKALLHQLCGVLPYSXDSFYLGRLGELNSLPLEVVGAMTIRV 180
Db 114 TGGGSGADRTEDATSLQKALLHQLCGVLPSSPDSFRLGRLGELNSLPLEVVGAMTIRV 173
QY 181 NSLTRGSHAVRLVYVLPALTNELNHGITTPIVLRGTISAGDLSPLSYAAITGHPDSKY 240
Db 174 NSLTRGSHAVRLVYVLPALTNELNHGITTPIVLRGTISAGDLSPLSYAAITGHPDSKY 233
QY 241 HYXHEGEXKIMXAREALFGLPEVYLGREGGLGNGTAVASASWATLALHDAMLSLS 300
Db 234 HYXHEGEXKILVAREAMALFNLPEVYLGREGGLGNGTAVASASWATLALHDAMLSLS 293
QY 301 QALTALVEAMVGHAGSFHPLHDVTRPHPTQIEVANIRITLLEGSKFAVHHEEVAKD 360
Db 294 QSLTAMTVAMVGHAGSFHPLHDVTRPHPTQIEVAGNIRKLLEGSPRAVHHEEQVKD 353
QY 361 DEGITLQDQYPRRTSPQWIGPLVSDMHAHAYLSLQAGOSTTNDNPLIDVENKXTHHGNF 420
Db 354 DEGITLQDQYPRRTSPQWIGPLVSDMHAHAYLTLEAGOSTTNDNPLIDVENKXTHHGNF 413
QY 421 QASAVXNTEKTRALALIGKLNFTQLEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 480
Db 414 QAAAVANTMEKTRRLGAGIGKLNFTQLEMLNAGNNGLPSCLAEDPSLSYHCKGLDIA 473
QY 481 AAAYTSELGHLNPTVTHVQPAKNGQAVNSLALISARRTAEANDVLSLLATHLYCVLQ 540
Db 474 AAAYTSELGHLNPTVTHVQPAKNGQAVNSLALISARRTAESNDVLSLLATHLYCVLQ 533
QY 541 AVDLRAMEPEFKQFPKXXXXLXOHFGXXXTXXXXXXELXXXKVKXLKRLQETNSYDL 600
Db 534 AIDLRAIEPEFKQFPKPALVSLIDQFSGAMTGSNLRDELVEKNKTLAKRLQETNSYDL 593
QY 601 EPRMHDASFAXGTGVVEXLSSXXXXXVSLAAVANAMKVAAXAEKALSTRXVXXFXAPSS 660
Db 594 VPRMHDASFAXGTGVVEXLSS--TSLSLAAVANAMKVAABESALSLTRQVRETFWASA 650
QY 661 SSPALXYLSRPTVXASFREELGVYARGGVFLGKQEVTTGSNYSRIYEAIKSGRINXV 720
Db 651 SSPALXYLSRPTQILYAFVREELGVYARGGVFLGKQEVTTGSNYSRIYEAIKSGRINNV 710
QY 721 LVHQLA 726
Db 711 LKQMLA 716

```

Search completed: September 9, 2004, 10:04:03
Job time : 92.3693 secs



Db 314 IGSFANVHEEAVKWDDEGILRDPYPLRTSPQWGLVSLDILAHAVLTIEAGSTTD 373

QY 404 NPLIDENKKTTHGSGNFQASAVXNTYKTRLMALALIGKLPOLTETMLNAGNTRGLPSCL 463

Db 374 NPLIDENKKTSHHGSGNFQAAVAANTWETKRLGLAIGKLPOLTETMLNAGNTRGLPSCL 433

QY 464 AAEDPSLSTYCKCLDIAAAAYTSELGHANPVTTHVQAPEMGNQAVNSIALISARTEA 523

Db 434 AAEDPSLSTYCKCLDIAAAAYTSELGHANPVTTHVQAPEMGNQAVNSIALISARTEA 493

QY 524 NDVLSLLATHLVYCVLQAVDLRAMEFEFKQEPXXXXLLXQHFQXXXXXXELXX 583

Db 494 NDVLSLLATHLVYCVLQAVDLRAIEFEFKQGPVAVSLIQHFGSAMTGSNTLDELVEK 553

QY 584 VXXKILKRLRQNTSYDLERWMDAFSXANGTVEKLSXXXXXVSLAAVNMKVALXKA 643

Db 554 VNKTLMKRLBQNTSYDLERWMDAFSPAAGTVEVLS--TSLSLAAVNMKVALXKAESA 610

QY 644 ISLTRVXXFXWAPBSSSPALXYLSPRIVLVSVREBELGVKARQGVFLGKQEVITGS 703

Db 611 ISLTRQVRETFMSPAASTSPALSYLSPRIQILYAVREBELGVKARQGVFLGKQEVITGS 670

QY 704 NVSRITAEAKSGRINXVLVKMLA 726

Db 671 NVSRITAEAKSGRINXVLVKMLA 693

5a

C/Species: Rhodotorula rubra
C/Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
C/Accession: S01997, S01990
R/Filipula, D.; Vasilat, C.A.; Levy, A.; Sykes, A.; Stransberg, R.L.
Nucleic Acids Res. 16, 11361, 1988
A/Title: Nucleotide sequence of gene for phenylalanine ammonia-lyase from Rhodotorula ;
A/Reference number: S01999; PMID:89083582; PMID:3205749
A/Accession: S01999
A/Molecule type: DNA
A/Residues: 1-713 <PIL>
A/Cross-references: EMBL:X13094; NID:925284; PIDN:CAA31486.1; PID:9259542
R/Vasilat, C.A.; Stransberg, R.L.; Sykes, A.; Levy, A.; Filipula, D.
Nucleic Acids Res. 16, 11382, 1988
A/Title: cDNA and genomic cloning of yeast phenylalanine ammonia-lyase genes reveal ger
A/Reference number: S01990; PMID:89083583; PMID:3205750
A/Accession: S01990

A;Residues: 1-674 <VAS>
A;Cross-references: EMBL:X13095

A; Introns: 110/2; 139/2; 372/3; 428/2; 482/3

C;Keywords: ammonia-lyase; carbon-nitrogen lyase
E:217-219/Cross-link; E:imide/none; (N1a Clv) #catalytic modified

F;218/Modified site: dehydroalanine (Ser) #status predicted

Query Match	86.8%;	Score 2922.5;	DB 2;	Length 713;
Best Local Similarity	82.0%;	Pred. No. 4.4e-222;		

matchines	55;	conservative	55;	mismatchines	65;	indeis	13;	gaps	5;
-----------	-----	--------------	-----	--------------	-----	--------	-----	------	----

[illegible][illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

121 TCECCSADPTEDATISLOKXIT EHOI CUII DEEYDGEYI CECI ENCI DI EYHECAMEYBY 100

Db 120 TGE GGSADPTREDAISLOKATLEHOTGGYTPTSM DGEALGRIENSTPIEYVEGAMTTRV 179

```

QY 181 NSLTGSHAVRLVLEALTNPLNGITPIVPLRGITISAGDLSPLSYIAAITGHPDSKY 240
DB 180 NSLTGSHAVRLVLEALTNPLNGITPIVPLRGITISAGDLSPLSYIAAITGHPDSKY 239
QY 241 HYXHEGEXKIXXAREALFGLPEVLGPKGGLVNGTAVASAMATLALHDAIMLSLS 300
DB 240 HY--DG--KIMSAGAILKGLQPVVLGPKGGLVNGTAVASAMATLALDAIVLSLLA 295
QY 301 QALITALVEAMGAGSHHPFLHDYTRPHPOIEVARNIRTLLEGSKFAVHHEEVVKVD 360
DB 296 QALITALVEAMGAGSHHPFLHDYTRPHPOIEVARNIRTLLEGSKFAVHHEEVVKVD 355
QY 361 DEGILRODYPPLRTSPQWLGPLVSDMIAHAVALSLGAGOSTTDNDPLIDVENKXTHGNGF 420
DB 356 DEGILRODYPPLRTSPQWLGPLVSDMIAHAVALSLGAGOSTTDNDPLIDVENKXTHGNGF 415
QY 421 QASAVXNTEKTRIALALIGKINFTQLTBMNAGNRLPSCLAEDPSLSYHCKGLDIA 480
DB 416 MASSVGNTEKTRIALALIGKINFTQLTBMNAGNRLPSCLAEDPSLSYHCKGLDIA 475
QY 481 AAAYTSELGHLNPTVTHQPAEMGNQAVNSIALISARRTEANDVLSILLATLHYCVLQ 540
DB 476 AAAYTSELGHLNPTVTHQPAEMGNQAVNSIALISARRTEANDVLSILLATLHYCVLQ 535
QY 541 AYDLRAMEFEFKQFXPXXXXLXQHPGXXXTXXXXXXELXXKXKXLEQTNSTYL 600
DB 536 AYDLRAMEFEFKQFXPXXXXLXQHPGXXXTXXXXXXELXXKXKXLEQTNSTYL 590
QY 601 EPPMHDAEXAAGTVYEXLSXXXXXVSLAANMKAAXAKAISTPRXVXXWMASS 660
DB 591 EPPMHDAEXAAGTVYEXLSXXXXXVSLAANMKAAXAKAISTPRXVXXWMASS 647
QY 661 SSPALXYLSPRTVLYSFVREELGVKARGDVFLLGCKQEVTTIGSNVSRITYEAIKSGRINXV 720
DB 648 SSPALXYLSPRTVLYSFVREELGVKARGDVFLLGCKQEVTTIGSNVSRITYEAIKSGCLAPV 707
QY 721 LVKMLA 726
DB 708 LVKMLA 713

RESULT 4
G84787
phenylalanine ammonia-lyase (PAL1) [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C/Accession: G84787
R/Lin. X.: Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Molnar, K.S.; Cronin, L.A.; Shen, M.; Varkken, S.E.; Umayam, L.; Tallon, L.;
Euse, D.; Nierman, W.C.; White, O.; Eiten, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; NCID:20083487; PMID:10617197
A/Accession: G84787
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-725 <STO>
A/Cross-references: GB:AE002093; NID:g6598547; P1DN:AA018156.2; GSPDB:GN00139
C/Genetics:
C/Superfamily: histidine ammonia-lyase

Query Match 28.8%; Score 369; DB 2; Length 725;
Best Local Similarity 35.5%; Pred. No. 3,3e-68;
Matches 258; Conservative 120; Mismatches 257; Indels 92; Gaps 22;

QY 40 GSXLPITYXQQLDVLVEXXLDXPELXDLGVSLTLGQVVGAAKRGKXRVXDSEIRX 99
DB 39 GAAAEQKSGHDEVKRWAB-FRKPVNVLGEBTLTGVAALSTIGSVKVELSETARA 97
QY 100 KIDKSVFLRXQLXNSV--YGVTTGFGSADRTEDTAISIQALLHQLCGVLPSTXDSF 157
DB 99 KIDKSVFLRXQLXNSV--YGVTTGFGSADRTEDTAISIQALLHQLCGVLPSTXDSF 157

```

```

DB 98 GYNASDWMESNMKGTDSYGVTTGFGATSHRTKNGVALQKELIRPLNAGIFGSKYET- 156
QY 158 XIGRLENSLLEEVKRGMTIRVNSLIRGSAVRLVLEALTNPLNGITPIVPLRGITIS 217
DB 157 -----SHTEHSATRAAMVLRINTLLQGFSGIRFELLEAITSLNNNTIPSLRGITIT 210
QY 218 ASGDSPLSYIAAITGHPDSKFAVHHEGEXKIXXAREALFGLPE--VLGPKGGLG 275
DB 211 ASGDLVPLSYIAAGLTGKPNKA--TGPNG--EALTAEARKLAGISSGFPDLOKREGAL 267
QY 276 VNGTAVASAMATLALHDAIMLSQALITALTVEAMGAGSHHPFLHDYTRPHPOIEV 335
DB 268 VNGTAVASAMATLALHDAIMLSQALITALTVEAMGAGSHHPFLHDYTRPHPOIEV 336
QY 336 ANINITLEGSKF-----AVHHEEVKVDDEGILRODYPPLRTSPQWLGPLVSDMIAHA 391
DB 327 AAIMHIDGSSYMLAQKHEMDPLQK-----KORVALRTSPQWLGQIE--VIRYA 379
QY 392 VLSLEAG--QSTTDNDPLIDVENKXTHGNGFQASAVXNTEKTRIALALIGKINFTQLT 450
DB 380 TKSIREINSVNDNDPLIDVSRNKAIHGNGFQGTPIGVMDNTRIAIAIGKIMRAQPSL 439
QY 451 LNAGNRLGIPCL--AAEDPSLSYHCKGLDIAAAVTSSELGHLNPTVTHQPAEMGNQAV 509
DB 440 VNDFYNNGLPSVLTASRNPSLDYGRKAELMASYCELOYLANPVTSHQSAEQHNDV 499
QY 510 NSIALISARRTEANDVLSILLATLHYCVLQAVDLRAMEFEFKQFXPXXXXLXQHPGX 569
DB 500 NSLGISSKRSSEAVDILKMTSTFLVALICQAVDLRHEBLRQ----- 543
QY 570 XTXXXXXXELXXKXKXLEQTNSTYL-----XKXKXKL--EQTNSTYDLPRMHAFAKAT 612
DB 544 --TVKNTVSQAVKXKLTGTGNGELHPSRFCEKDLTKVDRQVTVYADDP-----CSAT 595
QY 613 GTVEXLSXXXXXVSLAANA-----WKVAAEKALIS--LTRVYKXXFWXAPS 659
DB 596 YPLIQKLRQ---VYDHLINGESKNAVTSIFPKIGAFEBELKAVLPKEVABABAYDN 652
QY 660 SSSPALXYLSP-RTVLYSFVREELGVKARGDVFLLGCKQEVTTIGSNVSRITYEAIKSGRIN 718
DB 653 GTSALPNRIKESRYPYLFVREELGT-----ELITGKVTSPGEEDKVTALCAEKGII 707
QY 719 XVLVKKL 725
DB 708 DPMWEC 714

RESULT 5
S52990
phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Jul-1995
C/Accession: S52990; PQ0140
R/Wanner, L.A.; Li, G.; Ware, D.; Somsich, I.E.; Davis, K.R.
Plant Mol. Biol. 27, 327-338, 1995
A/Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
A/Reference number: S52990; NCID:9305677; PMID:7888622
A/Accession: S52990
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-725 <MAN>
A/Cross-references: EMBL:L33677; NID:g497418; P1DN:AA01870.1; P1D:g497419
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
R/Ohl, S.; Hedrick, S.A.; Chory, J.; Lamb, C.J.
Plant Cell 2, 837-848, 1990
A/Title: Functional properties of a phenylalanine ammonia-lyase promoter from Arabidopsis
A/Reference number: PQ0140; NCID:9305677; PMID:2152131
A/Accession: PQ0140
A/Molecule type: DNA
A/Residues: 1-240 <OH>
A/Cross-references: GB:X62747; NID:g16408; P1DN:CAA44609.1; P1D:g16409
A/Experimental source: Ecotype Columbia
C/Comment: Phenylalanine ammonia-lyase catalyzes the deamination of L-phenylalanine to L-

```

C:Genetics:
 A:Gene: PAL1
 A:Introns: 143/2
 C:Superfamily: histidine ammonia-lyase
 C:Keywords: ammonia-lyase; carbon-nitrogen lyase
 F:211-213/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:121/Modified site: dehydralanine (Ser) #status predicted

Query Match 28.7%; Score 965; DB 2; Length 725;
 Best Local Similarity 35.5%; Pred. No. 6,9e-68;
 Matches 258; Conservative 119; Mismatches 258; Indels 92; Gaps 21;

```

QY 40 GSXLPPTXXTQDLIVEXXLDADPTDXXELDYSILTGVDVGAARKGXVRYXSDDEIRX 99
DB 39 GAAAEQKMGSHLDEVKMAAE-FRKRVNVLGGSTLITGVAAIISTIGSVKLEBETARA 97
QY 100 KIDKSVFELRXQXLXNV--YGVTTGFGSADRTEDAIISQKALLEHQLCGVLPFXDSF 157
DB 98 GVNASDPMVMESNMKGTDYGVTTGFGATSHRTKXGVALQKELRFLNAGIFGSKET- 156
QY 158 KLGKLENSLPLEVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGITTVPVLRGTIS 217
DB 157 -----SHLPHSATRAAMLVRLNTLLQGSGRFELILAITSPNNITPSPURGIT 210
QY 218 ASGDLSPSTIAAATGHPDSKVYHHEGKIMKAREALFGLER--VVLGPRGGL 275
DB 211 ASGDLVPLSYIAGLITGRPSKA-TGPNG--EALTBEPFKLAGISSGFPLQPEGAL 267
QY 276 VNGTAVASAVATLALDHAMLSTLSQALTALTEAVNAGSFFHLDVTRPHTQIEV 335
DB 268 VNGTAVSGASAVLFEETNLSTVLAELIAVFAVWSGKP-EFTDLTRHLKHPGQIEA 326
QY 336 ARNRTLEGGXFP-----AVHHEEVKVKDDEGLRQDRYPLRTSPQWGLPYSDMIAHA 391
DB 327 AAVMEHLIDSSYMKLAQKLEHNDPLQKP-----KODRYALRTSPQWGLPQIE-VRIFA 379
QY 392 VLSLEAG-QSTTDNPLIDVENKXTHHGNFQASAVNXTMEKTRLALIGKLNFTOLTEN 450
DB 380 TKSIREINSVNDNPLIDVSRNKAHGNFQGPFGVSMNTRLAIRAIIGKLMFAQFSEL 439
QY 451 LNAAGNRGLPSCA-AEDPSLSYHCKGLDIAAAATSELGHLANVTTHQPAEMGNQAV 509
DB 440 VNPFYVNGPSTNLTARNSPLDGFGEAETIMASVSELQYLANVTSHVOSAQOHNDV 499
QY 510 NSLALISARRTAENDVLSLLATLHYCVLAQVLDLRAEFBEKQFPXKXXLLXQHFQX 569
DB 500 NSLGLISRSKTSADVILKLMSTTFLVAICQAVDLRHLLENLRQ----- 543
QY 570 KXTXXXXXXELXKV-----XCKLXKL--EQTNSYDLEPRMHAFAFXAT 612
DB 544 --TVKQTVSQVAKVLTGTGNGELHSPRCECDLLKVVDRQOVYIADDP-----CSAT 595
QY 613 GTVEXLSXXXXXXVSLAAVNA-----MKVAXAEKAIIS--LTRVYKXXFWXAPS 659
DB 596 YPLIQKLRQ---VIVDHALVNGSEKNATVSIFFHKGAFEEBELKAVLPKEVEAARAAYDN 652
QY 660 SSSPALXVLSLP-RTRVLYSFVRBELGVKARGDVFLEKQEVTTGSNVSRIEAIKSGRI 718
DB 653 GTSIAIPNRICEKRSYPLVRFVVEELGT-----ELLTGKVTSPGEEFPKVTALCEGKI 707
QY 719 XVLVKKL 725
DB 708 DPMKECL 714

```

RESULT 6
 S06475
 phenylalanine ammonia-lyase (EC 4.3.1.5) - rice
 C:Species: Oryza sativa (rice)
 C:Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 21-May-1999
 C:Accession: S06475; A33246; J00156
 R:Minami, E.; Ozeki, Y.; Matsuo, M.; Koizuka, N.; Tanaka, Y.
 Eur. J. Biochem. 185, 19-25, 1989

A>Title: Structure and some characterization of the gene for phenylalanine ammonia-lyase
 A:Reference number: J00156; KUID:90032676; PMID:2806257
 A:Accession: S06475
 A:Molecule type: DNA
 A:Residues: 1-701 <MIN>
 A:Cross-references: EMBL:116099
 A>Note: the authors translated the codon GTC for residue 66 as Leu, CAG for residue 35
 F:189-191/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F:190/Modified site: dehydralanine (Ser) #status predicted

Query Match 28.5%; Score 958.5; DB 2; Length 701;
 Best Local Similarity 36.0%; Pred. No. 2.1e-67;
 Matches 255; Conservative 116; Mismatches 266; Indels 71; Gaps 21;

```

QY 40 GSXLPPTXXTQDLIVEXXLDADPTDXXELDYSILTGVDV--VGAARKGXVRYXSDDEIR 97
DB 16 GAAAEQKMGSHLDEVKMAAQ-FREPLVKGATLRVQVAAVAAQADAAVAVELEDEEA 74
QY 98 EKXIDKSVFELRXQXLXNV--SVYGVTTGFGSADRTEDAIISQKALLEHQLCGVLPFXSD 155
DB 75 RPRVKAASWITTCIAHGDIDYGVTTGFGTSHRTKDGPLQVELLRHNAIGFGTSD 134
QY 156 SFYLGKLENSLPLEVRGAMTIRVNSLTRGSAVRLVLEALTNFLNHGITTVPVLRGT 215
DB 135 G-----HTLPSTRAAMLVRLNTLLQGSGRFELILAITSPNNITPSPURGIT 186
QY 216 ISAGDLSPSTIAAATGHPDSKVYHHEGKIMKAREALFGLER--VVLGPRGGL 273
DB 187 ITASGDLVPLSYIAGLITGRPSKA-TGPNG--EALTBEPFKLAGISSGFPLQPEGAL 243
QY 274 GLVNGTAVASAVATLALDHAMLSTLSQALTALTEAVNAGSFFHLDVTRPHTQIEV 333
DB 244 AIVNGTSVGSALAAVMDANILAVLSVLSAVCEVNNKPEYTHLTHKL-KHHNGSI 302
QY 334 EVARNIRTLBGSXAVHHEEVKVKDDEGL--RQDRYPLRTSPQWGLPYSDMIAHA 391
DB 303 DAAATMEHILAGSSPMSHAK--KXNEMDPLTKKQDRVALRNSPQWLGFOI-QVIRA-A 357
QY 392 VLSLEAG-QSTTDNPLIDVENKXTHHGNFQASAVNXTMEKTRLALIGKLNFTOLTEN 450
DB 358 TKSIREVNSVNDNPLIDVHGKALHGNFQGTIFGVSMNTRLAIRAIIGKLMFAQFSEL 417
QY 451 LNAAGNRGLPSCA-AEDPSLSYHCKGLDIAAAATSELGHLANVTTHQPAEMGNQAV 509
DB 418 VNEFYNNGLTSLNLAGSRPSIDYGFKEITIMASVSELQYLANPITNHQSAEQHNQDV 477
QY 510 NSLALISARRTAENDVLSLLATLHYCVLAQVLDLRAEFBEKQFPXKXXLLXQHFQX 569
DB 478 NSLGLVSAKRTLEADVILKLMSTTFLVAICQAVDLRHLLENKISSVKNCTQYAKVLTLM 537
QY 570 KXTXXXXXXELXKVXKXKLXRLQETNSYDLEP--RHHDAFSXATGTVVE-XLSXXXXXX 626
DB 538 NPTGLDSARPSKNTLLAIDR--EAVFSYADDPESANYPIMQKLRVAVLVEHALTSGDRA 596
QY 627 VSLAAVNAWKVAXAEKAIISLTRVYKXXFWXAPSSSSP-----ALXYLSL 670
DB 597 RGLRVLDHQQVGG-----APLCNAPGDRGRRRRRORTAPVANNRIVES 640
QY 671 RTRVLYSFVRBELGVKARGDVPF-LGKQEVTTGSNVSRIEAIKSGRI 717
DB 641 RSPFIYRVRBELG-----VFLTGKLTSPGEECKXVLTGISQKTL 682

```

RESULT 7
 T09777
 phenylalanine ammonia-lyase (EC 4.3.1.5) - loblolly pine

Figure 1

Species: *Pellioselinum crispum* (parsley,

2/20/2019 09:00:00 AM

Environ Biol Fish (2015) 98:1001–1011

1. The following are the names of the persons who have been appointed to the various positions in the Department of the Interior, Bureau of Land Management, for the year 1904:

A: Accession: 848726

[illegible]

EMDT - 2011 EC

EMPT Notes: 1004

[illegible][illegible][illegible]

M, CCCC. ENE

Key words: alliinase; alliinase-lyase; caldwell-1111

Entity	Actual	Predicted
Entity 1	1	1
Entity 2	1	1
Entity 3	1	1
Entity 4	1	1
Entity 5	1	1
Entity 6	1	1
Entity 7	1	1
Entity 8	1	1
Entity 9	1	1
Entity 10	1	1
Entity 11	1	1
Entity 12	1	1
Entity 13	1	1
Entity 14	1	1
Entity 15	1	1
Entity 16	1	1
Entity 17	1	1
Entity 18	1	1
Entity 19	1	1
Entity 20	1	1
Entity 21	1	1
Entity 22	1	1
Entity 23	1	1
Entity 24	1	1
Entity 25	1	1
Entity 26	1	1
Entity 27	1	1
Entity 28	1	1
Entity 29	1	1
Entity 30	1	1
Entity 31	1	1
Entity 32	1	1
Entity 33	1	1
Entity 34	1	1
Entity 35	1	1
Entity 36	1	1
Entity 37	1	1
Entity 38	1	1
Entity 39	1	1
Entity 40	1	1
Entity 41	1	1
Entity 42	1	1
Entity 43	1	1
Entity 44	1	1
Entity 45	1	1
Entity 46	1	1
Entity 47	1	1
Entity 48	1	1
Entity 49	1	1
Entity 50	1	1
Entity 51	1	1
Entity 52	1	1
Entity 53	1	1
Entity 54	1	1
Entity 55	1	1
Entity 56	1	1
Entity 57	1	1
Entity 58	1	1
Entity 59	1	1
Entity 60	1	1
Entity 61	1	1
Entity 62	1	1
Entity 63	1	1
Entity 64	1	1
Entity 65	1	1
Entity 66	1	1
Entity 67	1	1
Entity 68	1	1
Entity 69	1	1
Entity 70	1	1
Entity 71	1	1
Entity 72	1	1
Entity 73	1	1
Entity 74	1	1
Entity 75	1	1
Entity 76	1	1
Entity 77	1	1
Entity 78	1	1
Entity 79	1	1
Entity 80	1	1
Entity 81	1	1
Entity 82	1	1
Entity 83	1	1
Entity 84	1	1
Entity 85	1	1
Entity 86	1	1
Entity 87	1	1
Entity 88	1	1
Entity 89	1	1
Entity 90	1	1
Entity 91	1	1
Entity 92	1	1
Entity 93	1	1
Entity 94	1	1
Entity 95	1	1
Entity 96	1	1
Entity 97	1	1
Entity 98	1	1
Entity 99	1	1
Entity 100	1	1

quely match score 955.3; DB 2; length 118

Matches	264;	Conservative	114;	Mismatches	286;	Indels	87;	Gaps	21;
---------	------	--------------	------	------------	------	--------	-----	------	-----

14 NGAXNGAHAXXASXXXXXXXXXAGSALPTWXXIGLDLVEKXLADPAI.DDXXELDGS 7

DB 5 NGJING-HANGNGLDLCMIKEDPLNWGVAEALJGSHUDEVKRMVAE-YRKPVKLGGET 6;

74 LTLGbvG-AARkGRXVRvXDSDEIRKIDKSVFELRXOLNSV--YGVTTGFGGSADTR 118

63 LTISQVAISARDDSGVKVELSEEARAGVKASSDWMDSMKGTDSYGVTTGFGATSHRR 122

131 TEDAISLOCALLEHOLCGVPTSDSFXLGRGLE---NSLPLEVVBGAMTIRVNSLTBGH 18

123 TKGGALOEELIRLNAGI-----FGSGAEAGNNTT,PHSATBAAIT,VBINTT,I,OGV 1

188 SAVRLVLTAEALTNELNHGITPTVPVIBGTISASGDI.SPI.SYTAAATGHPNSKVHYXHEGX 24

174 SGTBEEETI.EATTKEI.NHNITBCT.PI.BGTTTASGDI.VPI.SYTAGI.ITGBBNSKA-YGPTGV 3

248 EXTMYAREATAT.EGI.EP--VYT.GPREGI.GI.YNGTAVSASMATT.AIHDAHMI.SIISOAITZ 30

233 --TISPEEAEKTAQVEGGEFFETDPKEGTAIYNGTAVGSCMAEMMTFEANNIAYITAEIMST36

306 I.TYERAMUGNAGSEHBEI NDUTPRVOTIEVABNTPIT BCSXZAVNNEFEFEYK/KRDECI 36

291 ТЕЛЕМОБИЛЬНЫЕ КОММУНИКАЦИИ И ИНТЕРНЕТ-ДОСТУП

366 BODENYI FESZEDŐT CÉLT TÖRTÉNT EGY KÖRÜLJÁRÓ ÉS A

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

1. The first part of the paper discusses the importance of the research and the objectives of the study.

[illegible]


```

Db      384  NPLIDVSRKALHGGNFGQGTPIGMSMONTRLAIAAGLWEAQFSEIIVNDVFNNGLPSNL 443
Qy      466  A-AEDPSLSYHKGGIDIAAAVYSELGHIANPVYTHQPAEMGNQAVNSALISARTE 522
Db      444  SGRNPSPIDYCKGKEIMASTCSLQFLAPVNVNHOSEQGNQDVNSLGLISSKTSIE 503
Qy      523  ANDVLSLLATHLYCVLQAVDLRAMEEFKKQEPXXXXLLXOHF-----GXXXTXXXX 577
Db      504  AVEIKLMTSTFVLVGLCAIDLRLHEENTKSYVNTKVSXAKVLTWGVNGELHPERFCE 563
Qy      578  XELXXXKXXX-----LKKLEQT-----NSVLEPRMMDASXATGYV 616
Db      564  KDLLEFVDEYVFAVITDPCSATYPLMOKRQCVLVEHALKNGSGNENMMNTSTFQKIATE 623
Qy      617  EXLSSXXXXXVSLAAVNAKVAAXEKAISLTRYVXXXFWXAPSSSSPALXYLSP----- 670
Db      624  DELK-----ALPKVEVSAR-----AALSGNPAL-----PRIIEC 655
Qy      671  RRRVLYSFYRELGIVKARGDVPLFGQEVITGVSVSHYIAISGRINNVKML 725
Db      656  KSYPLKFEVRKELGI-----EYLTGEKTSIGGEFFDVFLAMSKGELIIDLPLECL 705

```

RESULT 11

phenylalanine ammonia-lyase (EC 4.3.1.5) - parsley
C/Species: Petroselinum crispum (parsley)
C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-May-1999
C/Accession: S04463
R/Jois, R.; Dietrich, A.; Hahlbrock, K.; Schulz, W.
EMBO J. 8, 1641-1648, 1989
A>Title: A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification
A/Reference number: S04463; MUID:89356635; PMID:2767049
A/Accession: S04463
A/Molecule type: DNA
A/Residues: 1-716 <LOI>
A/Cross-references: EMBL:X15473
C/Genetics:
A/Introns: 136/2
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase, carbon-nitrogen lyase
F/202-204/Cross-link: 5-Imidazolinone (Ala-Gly) #status predicted
F/203/Modified site: dehydroalanine (Ser) #status predicted

Query Match	28.1%	Score 946.5;	DB 2;	Length 716;
Best local Similarity	36.7%	Pred. No. 1.9e-66;		
Matches 254;	Conservative 110;	Mismatches 286;	Indels 43;	Gaps 18;

RESULT 12

phenylalanine ammonia-lyase (EC 4.3.1.5) 1- soybean
C:Species: Glycine max (soybean)
C:Date: 06-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S22991
R:Frank, R.L.; Vodka, L.O.
DNA Seq. 1, 335-346, 1991
A:Title: Sequence and structure of a phenylalanine ammonia-lyase gene from Glycine max
A:Reference number: S22991; MUID:92190550; PMID:1799682
A:Accession: S22991
A:Molecule type: DNA
A:Residues: 1-713 <PRA>
A:Cross-references: EMBL:X52953; NID:G18376; PIDN:CAA37129.1; PID:G18377
C:Genetics:
A:Gene: PAL1
A:Introns: 11/2
A:Superfamily: histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase
F:189-201/Cross-link: 5-midazolone (Ala-Gly) #status predicted
/200/Modified site: denhydroalanine (Ser) #status predicted

QY 364 ILSDRYPLATSPQWIGPLVSDMIIHAVALSLDAG-QSTTDPLIDVENKXTHHGNGFOA 422
 Db 344 --KODRYALATSPQWIGPLVIE-VIRPSTKSIEREINSVNDPLIDVSRNKAHNGNGOG 399
 QY 423 SAVANTWEKTRIALALIGKLNFTQLTETMLNAGNRCGLPSC-L-AEDPSLSYHCKGLDIAA 481
 Db 400 TPIGVSMNDTRFLALASIGKLMFAQPSFELVNDPYNNGLPENLTASRNPSLDGFGKAEIAM 459
 QY 482 AATSEIGHLANPVTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATHLVYCVQA 541
 Db 460 ASYCELSQYLANPVTHVQPAEQNDVNSLIGLISRRKTEALIELIKMSSTFLLALCOA 519
 QY 542 VDIAPAEFFERKQFPXPKXXLLXQHFQXXXKXELXKXKXKXKRL-EQTNQSYD 599
 Db 520 IDLRHLEENLKNVKNVTVSQVSKR--ILTTGVNGELHPSRCEKDLKVDREYIFSYI 576
 QY 600 LEPAMDPAFSXATGTVEXISSXXXXVSLAANA-----KVAIAAEKAIIS--L 646
 Db 577 DDP-----CSATYPLMQKLRQ--VLVDHALVNAECEKDVNSIFQKIALFEEELKNL 627
 QY 647 TRYRXXFXKAPSSSPAL--XYLSPRTVLVSFVSEELGVKARQGVFLGKQEVYISGN 704
 Db 628 PKVEGA-RAAYSGRAAIENKIQEGRSYPLKFRFEEIGT-----GLITGEKVRSPGE 681
 QY 705 VSRIYEAIKSGRINXVLVKNL 725
 Db 682 FDKLFAMCGKIIDPLMECL 702

RESULT 13

S66343
 phenylalanine ammonia-lyase (EC 4.3.1.5) 1 - common tobacco
 C/Species: Nicotiana tabacum (common tobacco)
 C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 16-Jul-1999
 C/Accession: S66343
 R/Pukasa-Akeda, T.; Kung, S.; Watson, J.C.
 Plant Mol. Biol. 30, 711-722, 1996
 A/Title: Phenylalanine ammonia-lyase gene structure, expression, and evolution in Nicotiana
 A/Reference number: S66343; PMID:9619455; PMID:8624404
 A/Accession: S66343
 A/Status: preliminary
 A/Residue type: DNA
 A/Residues: 1-715 <PDB>
 A/Cross-references: EMBL:M64466; NID:G170349; PDB:AAA3122.1; PID:G170350
 C/Genetics:
 A/Gene: PAL1
 A/Introns: 133/2
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/201-Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/202/Modified site: dehydroalanine (Ser) #status predicted

Query Match 28.0%; Score 943; DB 2; Length 715;
 Best Local Similarity 36.4%; Pred. No. 3.7e-66;
 Matches 253; Conservative 107; Mismatches 254; Indels 82; Gaps 18;
 QY 68 ELQGYSLTLDVVGAA---RKGXRVVXSDERKXIDKSVFELKXQLKNSV--GYVTG 122
 Db 53 KLGESELTVAQVAIAARDKSAANGVAVLSEERAGVAKASDVMKMGKGTDSYGVTTG 112
 QY 123 FGGSADRTEDAIASLQKALIEHQLCGVLPFSXDSFYLGRLENS--LPLEVVRGAMTIRV 180
 Db 113 FGATSHRRRTKNGGLQGEILRFINAGV-----FGKGTESHTLPHSATRAAMTVRI 163
 QY 181 NSLTRGSAVRLVLELNTLFLNAGTTPVPLRGITISAGDLSPLSYIAAATGHPDSKV 240
 Db 164 NTLQGYSGIRFELLEITPLINSNTPLRGITITAGDVLPLSYIAGLTGHPNSKA 223
 QY 241 HVHKGEXKIMKAREALFGLRP--VLGPKEGGLGVGTAVASAMTALADHAMLSL 298
 Db 224 -VGNG--ELTINAEAFRAGVAGVGFELQPKEGGLAVNGTAVSGMASMVLFDENILAV 280

QY 299 LSCALTALTEAMVGHAGSFHPLHDVTRPHPTQIEVARNIRTLBEGSXPAVHEEYKV 358
 Db 281 MSEVLSAIFABVMNKP-EFTDHLJTHLKHHPGQIEAALIMEHILDDSSY-VKAAQCLHE 338
 QY 359 KODEGILRODRPLPSTSPQWIGPLVSDMIIHAVALSLDAGQSTTDNPLIDVENKXTHHG 418
 Db 339 MDPLQPKQDRYALATSPQWIGPLQI-EVIRATQIEREINSVNDNPLIDVSRNKAHNG 397
 QY 419 NFOASAVXWTEKTRIALALIGKLNFTQLTETMLNAGNRCGLPSC-L-AEDPSLSYHCKGL 477
 Db 398 NFQGTPIGSMNDTRFLALASIGKLMFAQPSFELVNDPYNNGLPENLTASRNPSLDGFGKA 457
 QY 478 DIAAAVYSELGHLANPVTHVQPAEMGNQAVNSLALISARRTAEANDVLSLLATHLVY 537
 Db 458 EIAAASYCELSQYLANPVTHVQPAEQNDVNSLIGLISARRKTEALIELIKMSSTFLLAL 517
 QY 538 VLOAVDLRAMEFEFRKQFPXPKXXLLXQHF-----GXXXKXKXKXKXKXKXKXKX 587
 Db 518 LCOALIDLRHLEENLKNVKNVTVSQVAKTLTMGANGELHPARFCKELALIVDREYIFAY 577
 QY 588 -----LXKRLDQ-----NSYDLEPRMDAFSXAATGTVEXISSXXXXVSLA 631
 Db 578 ADDPCSNVPLMQKLRQVLVDHANNGSEKVNNSIFQKIGAFED-----ELKA 627
 QY 632 VNAKVAIAEKALISLTVRXXFXKAPSSSPAL--XYLSPRTVLVSFVSEELGVAR 689
 Db 628 VLPKVESRAAL-----BSGPAIPNRTTECKRSYPLRFKRELGT---- 669
 QY 690 GDFVLGKQEVYISGNVSRIYEAIKSGRINXVLVKNL 725
 Db 670 -ELITGEKVRSPGECDKVFITAMCGIIDPLMECL 704

RESULT 14

T46172
 phenylalanine ammonia-lyase (EC 4.3.1.5) 2 [similarity] - Arabidopsis thaliana
 N/Alternate names: protein T4D2.190
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Mar-2000
 C/Accession: T46172; S52991
 R/Makatura, G.; Fartmann, B.; Dauner, D.; Seerr, W.; Holland, R.; Weichselgartner, M.
 submitted to the Protein Sequence Database, December 1999
 A/Reference number: 223025
 A/Accession: T46172
 A/Molecule type: DNA
 A/Residues: 1-717 <NYS>
 A/Cross-references: EMBL:AL12958; NID:G6434245; PDB:CA864229.1; PID:G6630746
 A/Experimental source: cultivar Columbia; BAC clone T4D2
 R/Wanner, L.A.; Li, G.; Ware, D.; Somsich, I.E.; Davis, K.R.
 Plant Mol. Biol. 27, 327-338, 1995
 A/Title: The phenylalanine ammonia-lyase gene family in Arabidopsis thaliana.
 A/Reference number: S52990; NID:95195160; PMID:7868622
 A/Accession: S52991
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Residues: 1-205, 'H', 207-234, 'E', 236-717 <MAN>
 A/Molecule type: DNA
 A/Residues: 1-205, 'H', 207-234, 'E', 236-717 <MAN>
 A/Cross-references: EMBL:U33678; NID:G497420; PDB:AAU18871.1; PID:G497421
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
 C/Genetics:
 A/Gene: PAL2
 A/Map position: 3
 A/Introns: 135/2
 A/Note: T4D2.190
 C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/203-Cross-link: 5-imidazolone (Ala-Gly) #status predicted
 F/204/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.9%; Score 939.5; DB 2; Length 717;
 Best Local Similarity 34.2%; Pred. No. 6.9e-66;
 Matches 251; Conservative 122; Mismatches 262; Indels 99; Gaps 20;
 QY 49 TQLDIVEXKLLADPKT-----DDXKE-----LDGYSLTLDVVGAA 84

```

Db      15  TKAAVTTKTLADPLNGLAADONKSGSHDEYVKWEEYRPPVNLGGELLTIGQVAIAST 74
QY      85  KGRXVRVXDSDELRXXIXDKSVEFLRXQLXNSV--YGVTTFGSGASDRTEDAISLQKALL 142
Db      75  VGSQVKVELAETSRAQVKAASDWVWESNMKGTDSGVTTTGFATSHRRTKGTALQGTBLI 134
QY      143  EHQLCGVLPSSXSFYLGKGLNLSLEVVNGAMTRVNSLFRGSAVRLVLEALINFL 202
Db      135  RFLNAGIFGNTKTC-----HTLPQSAIRAMVRRVNTLLQVSGIRFELLEIAITSL 187
QY      203  NHGITFVLPFGRTISASGDLSPSLYTAATITGHPDSKVNHXHEGEXIKMXAREALALFGL 262
Db      188  NNHISPSLPRGRTITSGDLVPLSTYAGILTRPNSKA-TGPDG--ESLTAKAEKXGI 244
QY      263  EP--VVLGPFGIGLVNGTAVSASMATLALHDAHMLSLISQALTALTYEAWGAGASHP 320
Db      245  STGFEPDLPQEGIALVNGTAVSGMAWMLFEANQAVLAELVLSAIFAEVNSGKP-EFTD 303
QY      321  FLHDVTRPHPTQTEVVRNIRTLLEGSGF-----AVHHEEVVKVODEGLRQDRPLRTS 375
Db      304  HLTHRLKHPQGLEAAIMEHILDGSSYKLAQKXHEMDPLQK-----KODRYALRTS 357
QY      376  PQWLGPLVSDMTHAAVLSLEAG-QSTTDNPLIDVENKXTHHGNFQASAVXNTWEKTRL 434
Db      358  PQWLGQPIE--VIRQATKSIEREINSVNDNPLIDVSRNKAIHGNFQGTPIGVSMNTRL 415
QY      435  ALALICKLNTQQTETMLNAGNRGLPSC--AAEDPSLSYHCKGLDIAAAATSEIGHLAN 493
Db      416  ALAIAIKLMPAQSELVNDFYNNGLPNSNLTAASSNSLDYGFEGABIAAASCSLQYLAN 475
QY      494  PVTTHVQAPEMGNOAVNSLALISARITAEANDVLSLTLATHLVQLQAVDIRAMEFEFK 553
Db      476  PVTSHVQSAHQNHQDONSIGLISRTSSEAVIDILKMGSTFLVGIQAVDIRHLEENLRQ 535
QY      554  QFAPXXXXLIXHFGKXXXIXXXXXXELKKV-----XXKLXRL--EQTN 596
Db      536  -----TVKNTVSOVAKKVLTTGNGELHPSRCEKDLKVVDRQVF 577
QY      597  SYDLAP--RWHDAFSXATGVVEVEXLSXXXXXVSLAAVAMKVAEKAIS--LTRXRX 652
Db      578  TYVDPCSAITPELMQRLQVIVDHALSNGETEKNAVTSIFQKIGAFEEELKAVLPKEVEA 637
QY      653  XFWXAPSSSPALXYLSP-RTRVLYSPVREELGVKARAGDVLGKQEVTTIGSNVSRIVEA 711
Db      638  ARAAYNGTAIPINRIKECRSYPLRFVREELGT-----LITGEKVSPSEEDKVFETA 692
QY      712  IKSGRINXVLVXKL 725
Db      693  WCEGKLIDPLMDCL 706

```

RESULT 15

T03663
 Phenylalanine ammonia-lyase (EC 4.3.1.5) - common tobacco
 C/Species: Nicotiana tabacum (Common tobacco)
 C/Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Jul-1999
 C/Accession: T03663
 A/Title: Phenylalanine ammonia-lyase in tobacco. Molecular cloning and gene expression of plant Physiol. 106, 877-886, 1994
 A/Reference number: Z14995; PMID:95125127; PMID:7824656
 A/Accession: T03663
 A/Status: Preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-712 <PEL>
 A/Cross-references: EMBL:X78269; NID:G633596; PIDN:CA55075.1; PID:9633597
 A/Experimental source: cultivar Samsum NN; tissue-type leaves after infection by tobacco
 C/Genetics:
 A/Gene: PAL
 C/Function:
 A/Description: catalyzes the deamination of L-phenylalanine to trans-cinnamic acid
 A/Note: strongly induced during the hypersensitive reaction to TMV or to a fungal elicitor

C/Superfamily: histidine ammonia-lyase
 C/Keywords: ammonia-lyase; carbon-nitrogen lyase
 F/198-200/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
 F/199/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.9%; Score 938; DB 2; Length 712;

Best Local Similarity 35.2%; Pred. No. 9e-66;

Matches 253; Conservative 115; Mismatches 260; Indels 90; Gaps 18;

```

QY      49  TQLDIYEXKLADPXTDDXXELDGYSLTLDGVYGAARK--GKXVRVXDSDELRXDKDVE 106
Db      33  SHLDEKKNVAA-FKRPVYKLGELLTYAQVALAKONAKTVKELSEGRARQVKAASD 91
QY      107  FLRXQLXNSV--YGVTTFGSGASDRTEDAISLQKALLEHQLCGVLPFSXDSFYLGRL 164
Db      92  VWDSMSKGTDSYGVTTGFGATSHRRTKGALQKELRFLNAGV-----FGNGTE 142
QY      165  --NSLPLEVVRGAMTRVNSLFRGSAVRLVLEALTNFLNHGTPPIVPLRGTISAGDL 222
Db      143  SCHTLPQSGTRAMVLRINTLLQVSGIRFELLEIAITLHNHNTPCPLPLGTTTASGDL 202
QY      223  PLSYTAATITGHPDSKVNHXHEGEXIKMXAREALALFGLP--VVLGPFGIGLVNGTA 280
Db      203  VPLSYTAGILTRPNSKA-IGPNC--ETLNAEAFRVAGVNSGFEELQPKGIALVNGTA 259
QY      281  VASASATLALHDAHMLSLISQALTALTYEAWGAGASHPFLHDVTRPHPTQIEVARNIR 340
Db      260  VSGGLSWVLPFANTLAAVSEVLSAIFAEVNSGKP-EFTDHLTHRLKHPQGLEAAIME 318
QY      341  TLLEGSPFAVHHEEVVKVODEGLRQDRYPLRTSPQWLGPLVSDMTHAAVLSLEAGQS 400
Db      319  HILDGSSY-VKAPQLKHEMDPLQKQDRYALRTSPQWLGQIEVINSATGMIEREINS 376
QY      401  TTDNPLIDVENKXTHHGNFQASAVXNTWEKTRLALIGKLNFTQLTETMLNAGNRGLP 460
Db      377  VNDNPLIDVSRNKALHGNFQGTPIGVSMNARFLASIGLMPAQSELVNDFYNNGLP 436
QY      461  SCP-AAEDPSLSYHCKGLDIAAAATSEIGHLANPVTTHVQAPEMGNOAVNSLISARR 519
Db      437  SNLTGRNNSLDYGRKGEIIMASVCSLQFLANPVNTHVQSAHQNHQDONSIGLISARK 496
QY      520  TAEANDVLSLTLATHLVQLQAVDIRAMEFEFKQFAPXXXXLIXHFGKXXXIXXXXXE 579
Db      497  TAEAVDILKLMSTYLVALQALDIRHLEENLRN-----A 531
QY      580  LXXKXKXKLXRLBQTNVYDLP-----RWHDAFSXATGVVEVEXLS 621
Db      532  VKNTVSOVAKKRTLTMGANGELHPSRCEKDLRVVDREYVFRVADACASANYPLMQKLRQ 591
QY      622  XXXXVYSLAAVAMKVAAXA-----EKAISLTRVYVXXFWXAPSSSPALXYLSP----- 670
Db      592  ---VLVDHALENGENEKNANSIFQKILAFEGELKAVLPKEVESARISLENGNPAIANRI 648
QY      671  ---RTRVLYSPVREELGVKARAGDVLGKQEVTTIGSNVSRIVEAIKSGRINXVLVXKL 725
Db      649  KCRCSPYPLRFVREELG-----AELLTGEKVRSPGEEDKVFETMNGQITDLSLECL 701

```

Search completed: September 9, 2004, 10:08:48
 Job time : 38.1203 secs

